

Robert Burduk
Marek Kurzyński
Michał Woźniak
Andrzej Żołnierek (Eds.)

Computer Recognition Systems 4

Advances in Intelligent and Soft Computing

95

Editor-in-Chief: J. Kacprzyk

Advances in Intelligent and Soft Computing

Editor-in-Chief

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

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

Vol. 84. Ryszard S. Choraś (Ed.)
*Image Processing and Communications
Challenges 2, 2010*
ISBN 978-3-642-16294-7

Vol. 85. Á. Herrero, E. Corchado,
C. Redondo, and Á. Alonso (Eds.)
*Computational Intelligence in Security
for Information Systems 2010*
ISBN 978-3-642-16625-9

Vol. 86. E. Mugellini, P.S. Szczepaniak,
M.C. Pettenati, and M. Sokhn (Eds.)
*Advances in Intelligent
Web Mastering – 3, 2011*
ISBN 978-3-642-18028-6

Vol. 87. E. Corchado, V. Snášel,
J. Sedano, A.E. Hassanien, J.L. Calvo,
and D. Ślęzak (Eds.)
*Soft Computing Models in Industrial and
Environmental Applications,
6th International Workshop SOCO 2011*
ISBN 978-3-642-19643-0

Vol. 88. Y. Demazeau, M. Pěchouček,
J.M. Corchado, and J.B. Pérez (Eds.)
*Advances on Practical Applications of Agents
and Multiagent Systems, 2011*
ISBN 978-3-642-19874-8

Vol. 89. J.B. Pérez, J.M. Corchado,
M.N. Moreno, V. Julián, P. Mathieu,
J. Canada-Bago, A. Ortega, and
A.F. Caballero (Eds.)
*Highlights in Practical Applications of Agents
and Multiagent Systems, 2011*
ISBN 978-3-642-19916-5

Vol. 90. J.M. Corchado, J.B. Pérez,
K. Hallenborg, P. Golinska, and
R. Corchuelo (Eds.)
*Trends in Practical Applications of Agents
and Multiagent Systems, 2011*
ISBN 978-3-642-19930-1

Vol. 91. A. Abraham, J.M. Corchado,
S.R. González, J.F. de Paz Santana (Eds.)
*International Symposium on Distributed
Computing and Artificial Intelligence, 2011*
ISBN 978-3-642-19933-2

Vol. 92. P. Novais, D. Preuveneers, and
J.M. Corchado (Eds.)
*Ambient Intelligence - Software and
Applications, 2011*
ISBN 978-3-642-19936-3

Vol. 93. M.P. Rocha, J.M. Corchado,
F. Fernández-Riverola, and A. Valencia (Eds.)
*5th International Conference on Practical
Applications of Computational Biology &
Bioinformatics 6-8th, 2011*
ISBN 978-3-642-19913-4

Vol. 94. J.M. Molina, J.R. Casar Corredera,
M.F. Cátedra Pérez, J. Ortega-García, and
A.M. Bernardos Barbolla (Eds.)
*User-Centric Technologies and
Applications, 2011*
ISBN 978-3-642-19907-3

Vol. 95. Robert Burduk, Marek Kurzyński,
Michał Woźniak, and Andrzej Żołnierak (Eds.)
Computer Recognition Systems 4, 2011
ISBN 978-3-642-20319-0

Robert Burduk, Marek Kurzyński,
Michał Woźniak, and Andrzej Żołnierek (Eds.)

Computer Recognition Systems 4

Editors

Dr. Robert Burduk
Wrocław University of Technology
Department of Systems and Computer
Networks, Wybrzeże Wyspiańskiego 27
50-370 Wrocław
Poland
E-mail: robert.burduk@pwr.wroc.pl

Prof. Marek Kurzyński
Wrocław University of Technology
Department of Systems and Computer
Networks, Wybrzeże Wyspiańskiego 27
50-370 Wrocław
Poland
E-mail: marek.kurzynski@pwr.wroc.pl

Dr. Michał Woźniak
Wrocław University of Technology
Department of Systems and Computer
Networks, Faculty of Electronics
Wybrzeże Wyspiańskiego 27
50-370 Wrocław
Poland
E-mail: michal.wozniak@pwr.wroc.pl

Dr. Andrzej Żołnierek
Wrocław University of Technology
Department of Systems and Computer
Networks, Wybrzeże Wyspiańskiego 27
50-370 Wrocław
Poland
E-mail: andrzej.zolnierek@pwr.wroc.pl

ISBN 978-3-642-20319-0

e-ISBN 978-3-642-20320-6

DOI 10.1007/978-3-642-20320-6

Advances in Intelligent and Soft Computing

ISSN 1867-5662

Library of Congress Control Number: 2011925360

©2011 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 to prosecution under the German Copyright Law.

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

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

Printed on acid-free paper

5 4 3 2 1 0

springer.com

Preface

Problem of pattern recognition is accompanying our whole life. We start to learn how to recognize simple objects like “dog”, “flower”, “car” when we are young and more sophisticated ones when we are growing up. Therefore the automatic pattern recognition is the focus of intense research of the Artificial Intelligence.

This book is the fourth edition of the monograph which focuses on the current directions in the modern compute pattern recognition. It offers actual review of the advances in pattern recognition and consists of 78 carefully selected works which have been reviewed carefully by the experts form the given domains. Chosen works were grouped into seven chapters on the basis of the main topics they dealt with:

1. “*Biometrics*” presents innovative theories, methodologies, and applications in the biometry.
2. “*Features, learning, and classifiers*” consists of the works concerning new classification and machine learning methods.
3. “*Image processing and computer vision*” is devoted to the problems of image processing and analysis.
4. “*Knowledge acquisition based on reasoning methods*” refers to the task of endowing expert system with the knowledge which could be used by its inference engine.
5. “*Medical applications*” presents chosen applications of intelligent methods into medical decision support software.
6. “*Miscellaneous applications*” describes several applications of the computer pattern recognition systems in the real decision problems as speech recognition, automatic text processing and analysis.

Editors would like to express their deep thanks to authors for their valuable submissions and all reviewers for their hard work. Especially we would like to thank Prof. Ewa Grabska and Prof. Katarzyna Stapor for their efforts to select articles for the chapter “*Knowledge acquisition based on reasoning methods*”, and Prof. Piotr Porwik who helped us to prepare the section devoted to Biometrics.

We believe that this book could be a great reference tool for scientists who deal with the problems of designing computer pattern recognition systems.

Wroclaw, 23rd May 2011

Robert Burduk
Marek Kurzyński
Michał Woźniak
Andrzej Żołnierak

Contents

Part I: BIOMETRICS

On-Line Signature Recognition Based on Reduced Set of Points	3
<i>Iwona Kostorz, Rafal Doroz</i>	
Pose Invariant Face Recognition Method Using Stereo Vision	13
<i>Jacek Komorowski, Przemyslaw Rokita</i>	
Subspace Algorithms for Face Verification	23
<i>Maciej Smiatacz</i>	
Fingerprint Orientation Field Enhancement	33
<i>Lukasz Wieclaw</i>	
Lip Print Recognition Based on Mean Differences Similarity Measure	41
<i>Lukasz Smacki and Krzysztof Wrobel</i>	
Human Vein Pattern Correlation – A Comparison of Segmentation Methods	51
<i>Rafał Kabaciński, Mateusz Kowalski</i>	
Knuckle Recognition for Human Identification	61
<i>Michał Choraś, Rafał Kozik</i>	

Part II: FEATURES, LEARNING AND CLASSIFIERS

Modified Ranked Order Adaptive Median Filter for Impulse Noise Removal	73
<i>Anna Fabijańska</i>	

Conceptual Design with the Use of Graph-Based Structures	83
<i>Grażyna Ślusarczyk</i>	
Estimation of Tolerance Relation on the Basis of Pairwise Comparisons	91
<i>Leszek Klukowski</i>	
Radial Basis Function Kernel Optimization for Pattern Classification	99
<i>Paweł Chudzian</i>	
Evaluation of Reliability of a Decision-Making Process Based on Pattern Recognition	109
<i>Artur Sierszeń, Lukasz Sturgulewski</i>	
Prototype Extraction of a Single-Class Area for the Condensed 1-NN Rule	119
<i>Marcin Raniszewski</i>	
Robust Nonparametric Regression with Output in $SO(3)$	127
<i>Grzegorz Jabłoński</i>	
Time Series Prediction with Periodic Kernels	137
<i>Marcin Michalak</i>	
Unified View of Decision Tree Learning Machines for the Purpose of Meta-Learning	147
<i>Krzysztof Grąbczewski</i>	
Designing Structured Sparse Dictionaries for Sparse Representation Modeling	157
<i>G. Tessitore, R. Prevede</i>	
Run-Time Performance Analysis of the Mixture of Experts Model	167
<i>Giuliano Armano, Nima Hatami</i>	
Fuzzy Clustering Finding Prototypes on Classes Boundary	177
<i>Michał Jezewski, Jacek Leski</i>	
Syntactic Pattern Recognition: Survey of Frontiers and Crucial Methodological Issues	187
<i>Mariusz Flasiński, Janusz Jurek</i>	
New Measures of Classifier Competence – Heuristics and Application to the Design of Multiple Classifier Systems	197
<i>Bartłomiej Antosik, Marek Kurzynski</i>	

Privacy Preserving Models of k-NN Algorithm	207
<i>Bartosz Krawczyk, Michal Wozniak</i>	
Some Properties of Binary Classifier with Fuzzy-Valued Loss Function	219
<i>Robert Burduk</i>	
Part III: IMAGE PROCESSING AND COMPUTER VISION	
Interactive Exploration of Image Collections	229
<i>Gerald Schaefer</i>	
Variable Complexity Reduction in Image Grade Decomposition	239
<i>Maria Grzegorek</i>	
Analysis of Inter-rater Agreement among Human Observers Who Judge Image Similarity	249
<i>Krzysztof Michalak, Bartłomiej Dzieńkowski, Elżbieta Hudyma, Michał Stanek</i>	
A Tool for Implementing and Evaluating Image Processing Algorithms inside FPGA Circuits	259
<i>E. Guzmán, I. García, M. Mendoza</i>	
Textural Features for Scribble-Based Image Colorization	269
<i>Michał Kawulok, Jolanta Kawulok, Bogdan Smolka</i>	
Palette Generation in $L^*a^*b^*$ Colour Space Using ΔE	279
<i>W.S. Mokrzycki and M. Tatol</i>	
Colour Image Quality Assessment Using the Combined Full-Reference Metric	287
<i>Krzysztof Okarma</i>	
Model-Based 3D Human Motion Capture Using Global-Local Particle Swarm Optimizations	297
<i>Tomasz Krzeszowski, Bogdan Kwolek, Konrad Wojciechowski</i>	
Gesture Data Modeling and Classification Based on Critical Points Approximation	307
<i>Michał Cholewa, Przemysław Głomb</i>	
Evaluation of Illumination Compensation Approaches for ELGBPHS	317
<i>Matthias Fischer, Marlies Rybnicek, Christoph Fischer</i>	

Efficient Stamps Classification by Means of Point Distance Histogram and Discrete Cosine Transform	327
<i>Paweł Forczmański, Dariusz Frejlichowski</i>	
Color Image Retrieval Based on Mixture Approximation and Color Region Matching	337
<i>Maria Luszczykiewicz-Piatek and Bogdan Smolka</i>	
LAMDA Methodology Applied to Image Vector Quantization	347
<i>E. Guzmán, J.G. Zambrano, I. García, Oleksiy Pogrebnyak</i>	
Detection of Semantically Significant Image Elements Using Neural Networks	357
<i>Jagoda Lazarek, Piotr S. Szczepaniak</i>	
The Effect of Multiple Training Sequences on HMM Classification of Motion Capture Gesture Data	365
<i>Michał Romaszewski, Przemysław Głomb</i>	
A Generic Approach to the Texture Detection Problem in Digital Images	375
<i>Christian Feinen, Marcin Grzegorzec, Detlev Droege, Dietrich Paulus</i>	
Waterball - Iterative Watershed Algorithm with Reduced Oversegmentation	385
<i>Michał Swiercz, Marcin Iwanowski</i>	
Kernel Postprocessing of Multispectral Images	395
<i>Marcin Michalak, Adam Świtoński</i>	
A Semi-local Topological Constraint for Efficient Detection of Near-Duplicate Image Fragments	403
<i>Mariusz Paradowski, Andrzej Śluzek</i>	
Morphology-Based Method for Reconstruction of Invisible Road Parts on Remote Sensing Imagery and Digitized Maps	411
<i>Bartłomiej Zielinski and Marcin Iwanowski</i>	
Spatio-Temporal Filters in Video Stream Processing	421
<i>Marek Szczepanski</i>	
Simplifying SURF Feature Descriptor to Achieve Real-Time Performance	431
<i>Marek Kraft, Adam Schmidt</i>	

Algorithm for Accurate Determination of Contact Angles in Vision System for High-Temperature Measurements of Metals and Alloys Surface Properties	441
<i>Tomasz Koszmider, Krzysztof Strzecha, Anna Fabijańska, Marcin Bakata</i>	

A Fuzzy C-Means Based Color Impulse Noise Detection and Its Benefits for Color Image Filtering	449
<i>Mihaela Cislariu, Mihaela Gordan, Victor Eugen Salca, Aurel Vlaicu</i>	

Stereovision System for Visually Impaired	459
<i>Rafal Kozik</i>	

Part IV: KNOWLEDGE ACQUISITION BASED ON REASONING METHODS

Artefactual Reasoning in a Hypergraph-Based CAD System	471
<i>Mehul Bhatt, Szymon Gajek, Ewa Grabska, Wojciech Palacz</i>	

A Shape Description Language for Osteophytes Detection in Upper Surfaces of the Metacarpophalangeal Joints	479
<i>Marzena Bielecka, Andrzej Bielecki, Mariusz Korkosz, Marek Skomorowski, Kamila Sprężak, Wadim Wojciechowski, Bartosz Zieliński</i>	

Using Graph Mining Approach to Automatic Reasoning in Design Support Systems	489
<i>Barbara Strug</i>	

A New Approach to Multi-class SVM-Based Classification Using Error Correcting Output Codes	499
<i>Wiesław Chmielnicki, Katarzyna Stapor</i>	

Fractal Dimension in the Landscape Change Estimation	507
<i>Piotr Łabędź, Agnieszka Ozimek</i>	

An Oversegmentation Method for Handwritten Character Segmentation	517
<i>Magdalena Brodowska</i>	

Part V: MEDICAL APPLICATIONS

Instantaneous Measurement of SNR in Electrocardiograms Based on Quasi-continuous Time-Scale Noise Modeling	529
<i>Piotr Augustyniak</i>	

Data Queuing for Non-uniform Telemedical Reporting	539
<i>Piotr Augustyniak</i>	
Analysis of Dispersion of Isopotential Areas Values in EEG Maps Sequences	549
<i>Hanna Goszczynska, Marek Doros, Leszek Kowalczyk, Paweł Hoser, Krystyna Kolebska</i>	
Robust Estimation of Respiratory Rate Based on Linear Regression	559
<i>J. Łęski, N. Henzel, A. Momot, M. Momot, J. Śledzik</i>	
Discriminating Power of an Sample Entropy and a Nonlinear Association Index in Prediction of a Preterm Labor Based on Electrohysterographical Signals	569
<i>Dariusz Radomski, Agnieszka Małkiewicz</i>	
Knowledge Extraction for Heart Image Segmentation	579
<i>Arkadiusz Tomczyk, Piotr S. Szczepaniak</i>	
Estimation of Wandering ECG Signal Baseline Using Detection of Characteristic Points in PQRST Complex for Avoiding Distortions in Complex Morphology	587
<i>Joanna Śledzik, Mirosława Stelengowska</i>	
Continuous Wavelet Transform as an Effective Tools for Detecting Motion Artifacts in Electrogastrographical Signals	597
<i>Barbara T. Mika, Ewaryst J. Tkacz, Paweł S. Kostka</i>	
Asymmetry of Digital Images Describing Melanocytic Skin Lesions	605
<i>Paweł Cudek, Jerzy W. Grzymala-Busse, Zdzisław S. Hippe</i>	
Fuzzy Clustering and Adaptive Thresholding Based Segmentation Method for Breast Cancer Diagnosis	613
<i>Paweł Filipczuk, Marek Kowal, Andrzej Obuchowicz</i>	
Application of Pattern Recognition Techniques for the Analysis of Histopathological Images	623
<i>Adam Krzyżak, Thomas Fevens, Mehdi Habibzadeh, Łukasz Jeleń</i>	
Part VI: MISCELLANEOUS APPLICATIONS	
Two-Stage Data Reduction for a SVM Classifier in a Face Recognition Algorithm Based on the Active Shape Model . . .	647
<i>Maciej Krol, Andrzej Florek</i>	

A New Star Identification Algorithm Based on Fuzzy Algorithms	657
<i>Shahin Sohrabi, Ali Asghar Beheshti Shirazi</i>	
User Authentication Based on Keystroke Dynamics Analysis	667
<i>Agata Kolakowska</i>	
Determining Feature Points for Classification of Vehicles	677
<i>Wieslaw Pamula</i>	
Edge Detection Using CNN for the Localization of Non-standard License Plate	685
<i>L.N.P. Bogavarapu, R.S. Vaddi, H.D. Vankayalapati, K.R. Anne</i>	
Artificial Neural Network Approach for Evaluation of Gas Sensor Array Responses	697
<i>Iwona Rutkowska, Barbara Flisowska-Wiercik, Przemysław M. Szczówka, Andrzej Szczurek</i>	
Developing Decision Tree for Identification of Critical Voltage Control Areas	707
<i>Robert A. Lis</i>	
A Review: Approach of Fuzzy Models Applications in Logistics	717
<i>Dragan Simić, Svetlana Simić</i>	
Expansion of Matching Pursuit Methodology for Anomaly Detection in Computer Networks	727
<i>Lukasz Saganowski, Tomasz Andrysiak, Michał Choraś, Rafał Renk</i>	
Stochastic Modelling of Sentence Semantics in Speech Recognition	737
<i>Włodzimierz Kasprzak, Paweł Przybysz</i>	
Thai Character Recognition Using “Snakecut” Technique	747
<i>Thakerng Wongsirichot, Piyawan Seekaew, Parunyu Arnon</i>	
Regularized NNLS Algorithms for Nonnegative Matrix Factorization with Application to Text Document Clustering	757
<i>Rafał Zdunek</i>	
Distant Co-occurrence Language Model for ASR in Loose Word Order Languages	767
<i>Jerzy Sas, Andrzej Zolnierek</i>	
Author Index	779

Part I
BIOMETRICS

On-Line Signature Recognition Based on Reduced Set of Points

Iwona Kostorz and Rafal Doroz

Abstract. In the paper two methods of signature points reduction are presented. The reduction is based on selecting signature's characteristic points. The first method is based on seeking points of the highest curvature using the IPAN99 algorithm. Parameters of the algorithm are selected automatically for each signature. The second method uses a comparative analysis of equal ranges of points in each signature. For both of methods the way of determination of characteristic points has been shown. As a result of experiments carried out the effectiveness of both methods and its usefulness for signature recognition and verification has been presented.

1 Introduction

Biometrics can be defined as the use of behavioural or physiological characteristics to recognize or verify the identity of an individual. Signature is one of the longest known security techniques. Pattern of signature has been an accepted form of credibility determination for many years (e.g. in the case of bank transactions).

Currently a lot of different approaches for signature verification have been proposed in the literature [1, 3, 5, 6, 10, 13]. Their actions are based on the calculation of distance such as Euclidean or Mahalanobis [6, 11, 14, 16]. They may also use different models of neural networks or Hidden Markov Models (HMM) [2, 8]. There are also methods based on an analysis of the signatures object characteristics (usually edges or angles) and reducing the problem to match these points [9].

The maximum size of the signature attributes vector representing the points obtained from a tablet can reach several thousand numbers. Analysing the entire received data string may be too time-consuming taking into the consideration the base of e.g. 10000 signatures. Therefore, the reduction of the vector size is done. Reduction of data can be achieved by identifying the various important points on contour or some attribute functions of a signature. In the presented work, the

Iwona Kostorz · Rafal Doroz

Institute of Computer Science, University of Silesia, ul. Bedzinska 39, 41-200 Sosnowiec
e-mail: {{iwona.kostorz, rafal.doroz}@us.edu.pl

characteristic points constitute a reduced subset of all points of a signature which were selected with the use of one of the methods presented in sections below. Those are the points representing e.g. the curves or corners of signature contour or equidistant points. The main problem is to identify such points, which can represent the shape of the object.

2 The Method for Determining the Characteristic Points

Signature points reduction is based on characteristic points determination. In this work two main approaches of signature points reduction were taken into consideration. The first one indicates the highest curvature points within the set of all signature points by means of modified IPAN99 algorithm [4]. In the second one all signature points are divided into equal ranges. Then from each range several points are selected. The method of points selection is fully described in the next section.

2.1 Modified IPAN99 Algorithm

Points with the highest curvature play an important role in perceiving the shapes by a human being. In the present paper, a very fast and efficient algorithm IPAN99 has been applied. This algorithm defines the corner as a point of a given curve, if we are able to inscribe in that curve a triangle with a given size and angle.

IPAN99 algorithm

In the first stage, the algorithm reviews the string of points and selects the candidates for corners. In the second stage, the redundant candidates are eliminated:

$$\begin{aligned} d_{\min}^2 &\leq |p - p^+|^2 \leq d_{\max}^2 \\ d_{\min}^2 &\leq |p - p^-|^2 \leq d_{\max}^2, \\ \alpha &\leq \alpha_{\max}, \end{aligned} \quad (1)$$

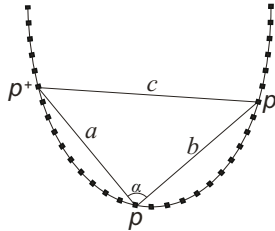


Fig. 1 Detection of points with the highest curvature by means of the IPAN99.

where:

$|p - p^+| = |a| = a$ - the distance between points p and p^+ ,

$|p - p^-| = |b| = b$ - the distance between points p and p^- ,

$\alpha \in [-\pi, \pi]$ - the triangle angle is defined as follows:

$$\alpha = \arccos \frac{a^2 + b^2 - c^2}{2ab}.$$

Values d_{min} and d_{max} are entered into the algorithm as its parameters (Fig. 2).

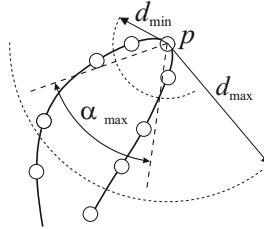


Fig. 2 Parameters of the IPAN99 algorithm.

A triangle which meets the conditions (I) is called an acceptable one. The searching for an acceptable triangle is started from the point p on the outside, that is, from the shortest length of the triangle sides, and stops if any part of the conditions (I) is fulfilled (therefore the restricted number of neighbouring points is taken into consideration). Among all acceptable triangles, the one with the smallest angle (the most acute) is selected $\alpha(p)$. The point p_i belongs to the neighbourhood of point p , if $|p - p_i|^2 \leq d_{min}^2$. IPAN99 algorithm parameters:

d_{min} - the minimal Euclidean distance from which neighbouring points will be considered to create triangles,

d_{max} - the maximal Euclidean distance to which neighbouring points will be considered to create triangles,

α_{max} - boundary angle specifying the minimal acuteness to classify the point as the candidate for the corner.

The candidate p point is rejected, if it has a shape neighbour that is p_i point, which is also a candidate, and which was assigned a greater strength of the corner: $\alpha(p) > \alpha(p_i)$.

Automatic selection of parameters of the IPAN99 algorithm

The method proposed in the paper for determining the characteristic points of the shape of a signature is based on two elements: an automatic selection of parameters of the IPAN99 algorithm and determination of additional points (so-called boundary points), which are not determined by the IPAN99 algorithm, but they are important for the signature analysis.

Signatures of different persons are characterized by different size and shape. This makes it impossible to determine a single universal set of parameter values of the IPAN99 algorithm for all signatures. This characteristic feature of the algorithm provided the basis for developing a method for automatic selection of a set of parameters of the IPAN99 algorithm. The selection is performed individually for each signature. It is based on the statistical analysis of the values of signature features - in this case: the shape, that is X and Y coordinates.

With the use of a tablet, a signature is represented by a set P of discrete points p of the signature:

$$P = \{p_1, p_2, \dots, p_n\} \quad (2)$$

where:

n - a number of signature points.

In order to calculate the statistical metrics, a set $L = \{l_{i,j}\}$ containing the Euclidean distances $l_{i,j}$ between successive points of the signature is determined. On the basis of the set L containing values of the distances between points of the signature chart, the mode M of the set L is determined [7].

Determination of additional characteristic points

In the developed method two sets of additional characteristic points are determined, which are not determined in the standard IPAN99 algorithm. The first one is a two-element set P_I , which contains the first and last point of the signature registered by the tablet:

$$P_I = \{p_1, p_n\} \quad (3)$$

The second set of characteristic points is determined on the basis of the set L . If the distance D_T between two points p_i and p_j is greater than the multiple of the value d_{max} :

$$\begin{aligned} |p_j - p_i| &> D_T \\ D_T &= c_3 \cdot d_{max} \end{aligned} \quad (4)$$

where:

c_3 - coefficient, both points are considered as additional characteristic points of the signature.

Figure 3 presents a chart of an example signature with marked characteristic points determined with the use of the IPAN99 algorithm and additional characteristic points.

Basing on the calculated value of the mode M [7], values of parameters of the IPAN99 algorithm are determined [7]:

$$d_{min} = c_1 \cdot M \quad (5)$$

$$d_{max} = c_2 \cdot M \quad (6)$$

$$D_T = c_3 \cdot d_{max} \quad (7)$$

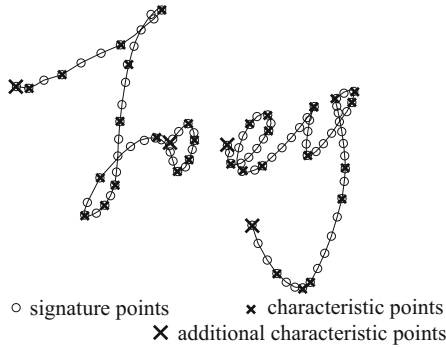


Fig. 3 Signature with marked characteristic points.

where:

c_1, c_2 - coefficients for the mode M ,

c_3 - coefficient for d_{max} .

2.2 Method of Ranges Analysis

Generally, the method of ranges analysis is based on dividing all signature's points into n equal ranges of points $R = (r_1, r_2, \dots, r_n)$, where $\forall_{r_i, r_j \in R} \text{card}(r_i) = \text{card}(r_j)$. The order of points is important at the line intersection. In the Fig. 4 the signature and two ranges of points (r_1 and r_2) are presented.

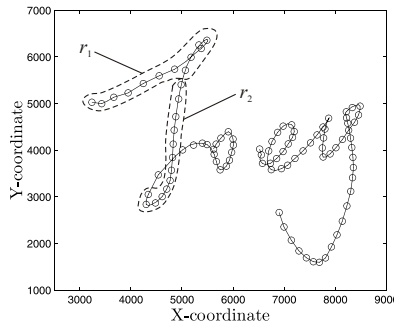


Fig. 4 The signature with two ranges (r_1, r_2) of points.

Characteristic points $CP = (cp_1, cp_2, \dots, cp_m)$ in each range were selected. Characteristic points were determined in two ways. The first way will be called MRA_E (Method of Ranges Analysis - Equidistant points). The second one will be called MRA_MM (Method of Ranges Analysis - Minimum Maximum value).

The MRA_E involves selecting a subset CP of the equidistant signature's characteristic points of all the signature's points (Fig. 5a). Equidistance of points is

determined by their indexes within the range of points (e.g. always the first, the second or the n -th point of the range is chosen). The number of elements of the subset CP was determined in two ways. In the first case, for each signature, the number of elements was adequate to the number of elements generated using the modified IPAN99 algorithm for particular signature. Basing on conducted studies [7] on database of signatures using modified IPAN99 algorithm, the mode value $M = 40$ of sizes of obtained sets of signatures' characteristic points was determined. Hence, in the second case number of elements ranges from 0.5 to 1.5 of mode M of number of all characteristic points determined using modified IPAN99 algorithm. In the MRA_MM, it was assumed that the number of elements of reduced subset CP of characteristic points:

- should range from 0.5 to 1.5 of mode M of number of all characteristic points determined using modified IPAN99 algorithm,
- only two points should be selected within each range,
- selected points ought to meet the condition that their Y coordinates reached only extremal values within each range (minimum and maximum) (Fig. 5b).

In the same way another subset of characteristic points was determined. This time, points which X coordinates reached minimum and maximum values within the range were selected. These two subsets were analysed separately.

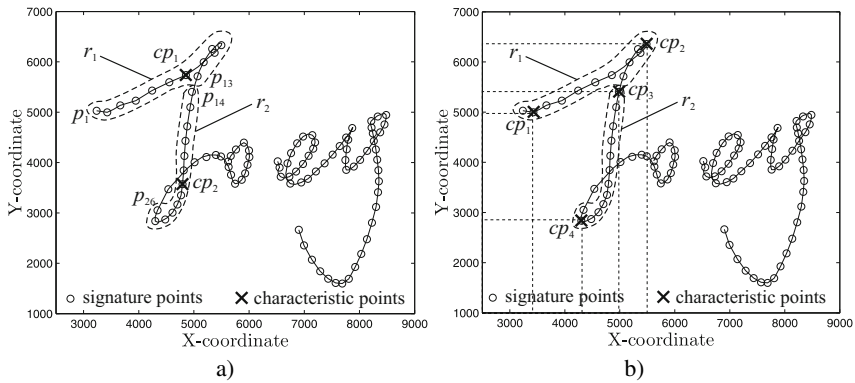


Fig. 5 The signature with: a) equidistant points, b) minimum and maximum values within each range.

3 Tests

The studies aimed at determining the effectiveness of the signature verification based on a reduced set of signature points. The set of the analysed signatures contained 96 signatures coming from 16 persons. The signatures were taken from the SVC2004 database (skilled forgeries). Each of the persons put 4 original signatures, while the number of false signatures for each person was 2. The values of the similarity between individual signatures were calculated with the use of the ratio R^2 [6].

From the conducted studies [7] it appears that the smallest EER=4.14% was obtained for the following combination of parameters: $c_1 = 1, c_2 = 2, c_3 = 2$. In the next stage of research the proposed modified IPAN99 algorithm method [7] was compared with other methods of signature's set of points' reduction. The results of effectiveness of proposed method of ranges analysis are presented on bar charts below (Fig.6).

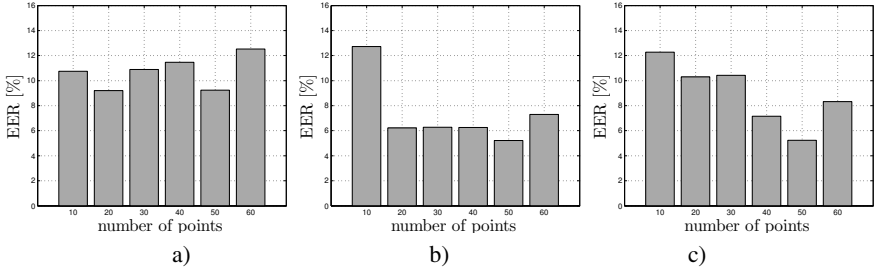


Fig. 6 Research results for a) MRA_E method, b) MRA_MM_X method, c) MRA_MM_Y method.

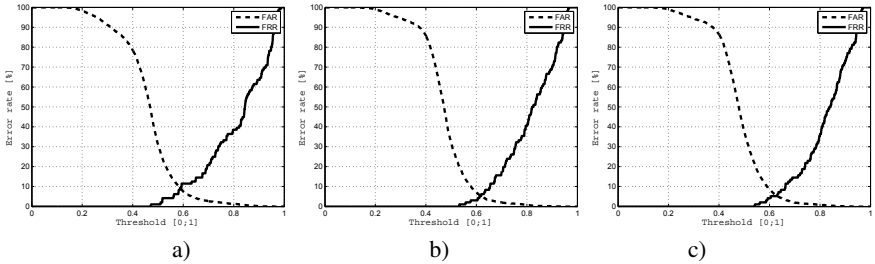


Fig. 7 FAR/FRR plots for best results of a) MRA_E method, b) MRA_MM_X method, c) MRA_MM_Y method.

The charts provide information about the EER values depending on assumed number of characteristic points. It can be seen from the charts that the EER value reaches a low point for sets of characteristic points which sizes were about 50.

The results of both proposed methods of generating subsets of signature's characteristic points were compared with the results of methods presented by its authors in the referenced paper [15]. The comparison of EER values is shown in the table [1]. Additionally, Figure 7 presents FAR/FRR plots for best results obtained for Method of Ranges Analysis.

Table 1 The comparison of ERR values for proposed methods and methods referenced in the literature.

	EER [%]
min MRA_E	9.23
min MRA_MM_X	5.21
min MRA_MM_Y	5.24
modified IPAN99	4.14
random [15]	7.10
genetic algorithm [15]	7.70

4 Conclusions

In the paper two methods of signature points reduction were presented. Both of them were tested on the same database of signatures. The studies showed that it was possible to recognize signatures basing on an analysis of only the characteristic points.

As a result of carried out investigation the lowest value of EER= 4.14% was obtained with the use of modified IPAN99 method. In comparison with effectiveness of method of range analysis it's more effective then proposed MRA_MM (Method of Ranges Analysis - Minimum Maximum value. By contrast, it's twice more effective then MRA_E (Method of Ranges Analysis - Equidistant points).

Moreover, the results of the other authors' methods [15] (random and genetic algorithm) that two proposed methods were compared with are significantly less effective then modified IPAN99 algorithm (their ERR values range from 7% to 8%) .

References

- [1] Al-Shoshan, A.I.: Handwritten Signature Verification Using Image Invariants and Dynamic Features. In: International Conference on Computer Graphics, Imaging and Visualisation. pp. 173–176 (2006)
- [2] Bajaj, R., Chaudhury, S.: Signature Verification Using Multiple Neural Classifiers. Pattern Recognition 30(1), 1–7 (1997)
- [3] Bolle, R.M., Pankanti, S., Ratha, N.K., Connell, J.H.: Guide To Biometrics, p. 75. Springer, New York (2009)
- [4] Chetverikov, D., Szabo, Z.: Detection of high curvature points in plannar curves. In: 23rd Workshop of the Austrain Pattern Recognition Group, pp. 175–184 (1999)
- [5] Doroz, R., Porwik, P., Para, T., Wrobel, K.: Dynamic signature recognition based on velocity changes of some features. International Journal of Biometrics 1(1), 47–62 (2008)
- [6] Doroz, R., Wrobel, K.: Method of Signature Recognition with the Use of the Mean Differences. In: Proceedings of the 31st International IEEE Conference on Information Technology Interfaces. Croatia, pp. 231–235 (2009)
- [7] Doroz, R., Kistorz, I.: The method for determining the characteristic points of Signatures based on IPAN99 algorithm. Journal of Medical Informatics and Technologies 15, 41–46 (2010)
- [8] El Yacoubi, A., Gilloux, M., Sabourin, R., Suen, C.: Unconstrained Hand Written Word Recognition using Hidden Markov Models. IEEE Transation on Pattern Analysis and Machine Intelligence 2(8), 752–760 (1999)

- [9] Feng, H., Wah, C.H.: Online signature verification using a new extreme points warping technique Source. *Pattern Recognition Letters* 24, 2943–2951 (2003)
- [10] Impedovo, S., Pirlo, G.: Verification of Handwritten Signatures: an Overview. In: 14th International Conference on Image Analysis and Processing, pp. 191–196 (2007)
- [11] Kato, N., Abe, M., Nemoto, Y.: A Handwritten Character Recognition System Using Modified Mahalanobis Distance. *Trans. IEICE J79–D–II(1)*, 45–52 (1996)
- [12] Piyush Shankera, A., Rajagopalan, A.N.: Off-line signature verification using DTW. *Pattern Recognition Letters* 28, 1407–1414 (2007)
- [13] Porwik, P.: The Compact Three Stages Method of the Signature Recognition. In: 6th International IEEE Conference on Computer Information Systems and Industrial Management Applications, pp. 282–287 (2007)
- [14] Schimke, S., Vielhauer, C., Dittmann, J.: Using Adapted Levenshtein Distance for On-Line Signature Authentication. In: *Proceedings of the 17th International Conference*, vol. 2, pp. 931–934 (2004)
- [15] Wirotius, M., Ramel, J., Vincent, N.: Selection of Points for On-Line Signature Comparison. In: *International Workshop on Frontiers in Handwriting Recognition*, Los Alamitos, CA, US, pp.1550–5235 (2004)
- [16] Wrobel, K., Doroz, R.: The new method of signature recognition based on least squares contour alignment. In: *Proceedings of the International IEEE Conference on Biometrics and Kansei Engineering*, pp. 80–83 (2009)

Pose Invariant Face Recognition Method Using Stereo Vision

Jacek Komorowski and Przemyslaw Rokita

Abstract. The paper presents a method to decrease dependency of image-based (appearance-based) face recognition solutions, such as Eigenfaces method, on a head orientation. Using information of a human head shape, recovered by stereo vision technique, a normalized face image (*en face* view) is synthesized. The normalized image is then processed by an image-based recognition subsystem. To effectively deal with large head rotations, where large area of the face is invisible due to occlusion, Eigenfaces recognition method is extended to operate on half-face images. Presented technique is verified experimentally to prove that it can be used to increase performance of image-based face recognition methods in applications where head pose of the subject is not controlled.

1 Introduction

Face recognition for biometric purposes is a field of a great practical importance and active development [11]. However classical image-based (appearance-based) recognition methods, such as Eigenfaces [8] or Fisherfaces, are very dependent on head orientation [10]. This is because 2D images of 3D face objects change significantly due to viewpoint variations.

Face recognition surveys [10][11] acknowledge performance degradation of image-based recognition systems when head orientation is not controlled but no quantitative results are given. To have an idea on this dependency in Experiment 2 we estimate an impact of head rotation on performance of Eigenfaces method. It can be seen that Euclidean distance between feature vectors of two images of the same face, one *en face* view and the other rotated out-of-plane in the left-right direction,

Jacek Komorowski

Maria Curie-Skłodowska University, Institute of Computer Science, Lublin, Poland
e-mail: jacek.komorowski@gmail.com

Przemyslaw Rokita

Warsaw University of Technology, Institute of Computer Science, Warsaw, Poland
e-mail: p.rokita@ii.pw.edu.pl

increases significantly even for small rotation angles. The greater is the distance, the bigger chance for false recognition. In our test scenario yaw angles above 18° yielded incorrect identification but this number is dependant on gallery size and content.

Advanced face recognition techniques, such as fitting to a 3D morphable model [11], overcome dependency on head orientation by using a morphable model of a human face. The model is fit to the face image being recognized and fitting parameters are used as features in subsequent recognition phase. In our research we explored a different approach without a need for encoding a morphable head model into a recognition system. Instead information of a human face shape is recovered using stereo vision technique during system operation.

2 Pose Invariant Head Recognition Method

The idea of our method is presented on Fig. 1. Gallery of the subjects in the recognition system consists of normalized, frontal view, face images, taken in neutral lighting conditions (see Fig. 3(a)). For identification, a pair of stereo images, taken with uncontrolled orientation, is used. So the solution may be deployed in situations where subject cooperation during identification is not required, such as scanning for faces of known suspects in public places like airports.

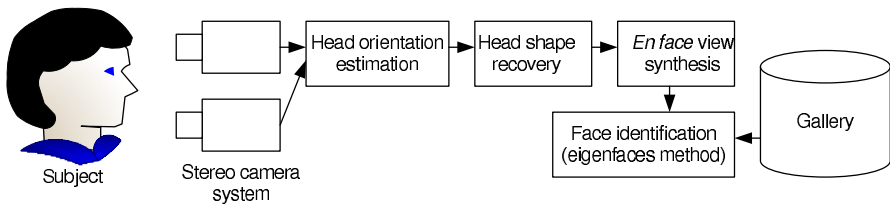


Fig. 1 Recognition system concept

Main steps of the subject identification in our method are as follows:

1. Acquisition of a pair of stereo images of the subject
2. Estimation of a head orientation
3. Recovery of a head shape using stereo vision technique
4. Synthesis of a normalized face image (*en face* view)
5. Subject identification using an image-based recognition method. In our implementation Eigenfaces algorithm was used but other appearance-based recognition method can be deployed.

The stereo camera system must be initially calibrated in order to estimate its intrinsic and extrinsic parameters. This is done using using OpenCV [12] library implementation of Zhang algorithm [12]. Using chessboard like calibration pattern, intrinsic

¹ Available at <http://opencv.willowgarage.com/wiki/>

and extrinsic parameters of the stereo rig are computed. In the subsequent steps the following parameters computed during calibration procedure are used: left and right camera 3x3 intrinsic matrices (K_l^* , K_r^*), left and right camera distortion coefficient vectors (d_l , d_r), 3x3 rotation matrix (R_r) and 3x1 translation vector (T_r) relating left and right cameras.

Head orientation estimation

Estimation of the head orientation can be done using one of the published methods for fine head pose estimation, e.g. [9]. This method is reported to estimate horizontal head orientation with 8.5% accuracy and vertical orientation with 12.5% accuracy. The outcome of the head pose estimation algorithm are three rotation angles: yaw angle φ_X (right to left rotation of the head), pitch angle φ_Y (up and down movement of the head), and roll angle φ_Z (right to left bending of the neck). Currently we haven't implemented an automatic head pose estimation method and during our experiments rotation angles were estimated manually.

Head shape recovery and en face view synthesis

To recover shape of the head a technique described by us in [4] is used. An input consists of a pair of head images from a stereo camera set. Firstly images are rectified using stereo rig intrinsic (K_l^* , K_r^* , d_l , d_r) and extrinsic (R_r , T_r) parameters estimated during an initial calibration. The result is a pair of rectified images and intrinsic matrices of a left and right rectified camera (K_l and K_r respectively). Then segmentation is done to produce a set of pixel coordinates $\mathcal{F} \subseteq \mathbb{N} \times \mathbb{N}$ constituting a region of the reference image containing a face. As subjects are placed on a neutral background a simple technique based on thresholding is sufficient.

Let (X, Y, Z) be a scene point in a camera coordinate system and (x, y) coordinates of its projection onto a camera image plane. Depth of a pixel (x, y) is defined as a Z-coordinate of a corresponding scene point (X, Y, Z) , that is $depth(x, y) = Z$. Depth is inversely proportional to a disparity, which is defined as a difference between coordinates of the projections of the scene point onto left and right image planes. Assuming left and right cameras have the same focal length f and distance between their principal points is T_r depth can be calculated using the formula:

$$depth(x, y) = \frac{f \|T_r\|}{disp(x, y)} . \quad (1)$$

Semi-global matching using mutual information algorithm [3] is used to compute a disparity map $disp(x, y)$ for each pixel (x, y) of a reference image contained within a face region \mathcal{F} .

Depth-image-based rendering technique [2] is used to synthesize a normalized frontal view of the subject face. A virtual view of a scene from a novel view point is synthesized using a reference image with associated per-pixel depth information. First, original image points are reprojected into 3D space, using per-pixel depth data.

Then 3D points are projected onto an image plane of a virtual camera, located at the required viewing position. Let $\mathbf{M} = (X, Y, Z, 1)^T$ be 3D homogeneous coordinates of the scene point and $\mathbf{m}_1 = (x, y, 1)$ it's projection onto reference camera image plane. \mathbf{M} and \mathbf{m}_1 are related by the camera perspective projection equation [6]:

$$z_1 \mathbf{m}_1 = K_1 [I|0] \mathbf{M} \quad (2)$$

where z_1 is the scene point depth, K_1 is a 3x3 rectified reference camera intrinsic matrix and $[I|0]$ is 4x3 matrix consisting of 3x3 identity matrix and 3x1 zero vector. Rearranging (2) we can derive 3D coordinates of a scene point \mathbf{M} :

$$\mathbf{M} = z_1 (K_1)^{-1} \mathbf{m}_1 \quad (3)$$

Let R_v be 3x3 rotation matrix and T_v 3x1 translation vector defining the required position of a virtual camera relative to the reference camera. Projection \mathbf{m}_v of a point $\mathbf{M} = (X, Y, Z, 1)^T$ onto a virtual camera image plane satisfies following perspective projection equation:

$$z_v \mathbf{m}_v = K_v [R_v | T_v] \mathbf{M} \quad (4)$$

Assuming that virtual camera intrinsic matrix K_v is the same as the reference camera intrinsic matrix K_1 and substituting (3) into (4) we obtain:

$$z_v \mathbf{m}_v = z_1 K_1 R_v (K_1)^{-1} \mathbf{m}_1 + K_v T_v \quad (5)$$

Depth from a virtual camera z_v can be assumed arbitrary as homogeneous coordinates scaled by a constant factor identify the same point and depth from a reference camera $z_1 = \text{depth}(x, y)$. So finally we get a formula:

$$z_v \mathbf{m}_v = \text{depth}(x, y) K_1 R_v (K_1)^{-1} (x, y, 1)^T + K_v T_v \quad (6)$$

Applying (6) for a point $(x, y, 1)^T$ from reference image we can get a point \mathbf{m}_v on the synthesized image it's mapped to. Repeating this procedure for all points from a segmented face region \mathcal{F} on a reference image produces an image from a novel viewpoint. As more than one pixel from a reference image may be mapped to the same pixel in a synthesized image a visibility problem must be solved. This is done using a simple z-buffering technique. Finally to deal with resampling issues a post processing step is necessary. When mapping pixels from a reference image some pixels in a synthesized image are not assigned any value. Telea inpainting algorithm [7] is used to fill uninitialized pixels within a face area on a synthesized image.

The final results are depicted on Fig. 2. Fig. 2(a) shows 3 examples of input images (reference images from stereo pairs) and Fig. 2(b) depicts results of view normalization on each input image.

Another problem is an occlusion as part of the face is not visible on both input images and pixel intensity values for these areas are unknown. Small occluded areas are filled using inpainting but the problem becomes more serious as yaw rotation angle increases. To deal with this difficulty Eigenfaces recognition algorithm was adapted to use half-face images when yaw angle is large.



Fig. 2 Normalized *en face* view synthesis results

Recognition

Normalized *en face* face image is converted into grey scale and fed into an image-based recognition subsystem. For our implementation we adapted a well-known Eigenfaces method [8], based on principal component analyses technique. The method consists of 2 steps: training and identification. Face images are represented as vectors x_i created by concatenation of image rows. Each image x_i is an element of high dimensional space \mathbb{R}^D , where D equals a number of pixels.

Initially an Eigenfaces recognition subsystem must be trained to learn a basis for a low dimensional representation of faces. Training is done using sufficiently large (usually from a few hundreds to a few thousands) gallery of normalized face images $\{x_i\}$. An average face \bar{x} is calculated as $\bar{x} = \sum_{i=1}^N x_i / N$, where N is a number of gallery images. For each face image x_i a deviation from a mean face \tilde{x}_i , referred as a *caricature*, is defined as $\tilde{x}_i = x_i - \bar{x}$.

Principal component analyses procedure is applied to the vector set $\{\tilde{x}_i\}$. It produces a set $\{u_n\}$ of $N - 1$ orthonormal vectors which best describe distribution of the data. The k -th vector u_k is chosen such that

$$\lambda_k = \frac{1}{N} \sum_{n=1}^M (u_k^T \tilde{x}_n)^2$$

is maximum, subject to

$$u_k^T u_l = \delta_{lk} = \begin{cases} 1 & \text{if } l = k \\ 0 & \text{otherwise} \end{cases}$$

Vectors u_k and scalars λ_k are eigenvectors and eigenvalues of covariance matrix

$$C = \frac{1}{M} \sum_{n=1}^N \tilde{x}_n \tilde{x}_n^t = AA^T$$

where matrix $A = [\tilde{x}_1, \dots, \tilde{x}_N]$. Dimensionality of AA^T is D^2 , so for a typical image size direct eigenvalue decomposition is not computationally feasible (e.g. for 128...128 image $AA^T \in \mathbb{R}^{16384}$). But usually a number of training data points N

is much lower than D and AA^T has only $N - 1 \ll D^2$ non-zero eigenvectors. A computation of u_k is efficiently done by first calculating $N - 1$ eigenvectors of $A^T A$ matrix, which has dimensionality $N \times N$. These eigenvectors are denoted as v_k . Then non-zero eigenvectors of AA^T are computed using the formula

$$u_l = \sum_{k=1}^N v_{lk} \tilde{x}_k$$

where v_{lk} is k -th element of a vector v_l .

Vectors in $\{u_n\}$, referred as *eigenfaces*, span a low dimensional space called *face space*. Projection of high-dimensional face caricatures \tilde{x} onto a *face space* produce a low dimensional face representation $\phi(x) \in \mathbb{R}^k$, which is used as a feature vector in subsequent recognition stage. j -th element of a feature vector $\phi(x)$ is calculated as $\phi^j(x_i) = \langle \tilde{x}_i, u_j \rangle$.

During identification step a feature vector $\phi(y)$ corresponding to a normalized input face image y is calculated by projecting caricature $\tilde{y} = y - \bar{x}$ onto a *face space*. Then similarity to each image in the gallery is evaluated using as a measure an Euclidean distance between corresponding feature vectors $\mathbf{dist}(x, y) = \|\phi(x) - \phi(y)\|_2$. A gallery image closest to the input image is chosen as a match.

Half-faces for recognition

The main problem we have to deal with is the fact, that even with a moderate yaw rotation angle a part of the face is occluded and not visible on any of the two cameras. It can potentially be solved by using face symmetry assumption and recovering missing pixel intensities using known values from the other side of the face. This would require an exact estimation of a face symmetry axis which can be difficult if large part of the face is occluded.

We propose a simpler approach to use for recognition only a fully visible half of the face image. To implement this idea three different sets of eigenfaces are computed using training gallery images. One is based on full face images, second on right half-face images and third on a left half-face images. Examples are depicted on Fig 3. Then three feature vectors are associated with each gallery image. One is computed using full image and set of eigenfaces based on full face images, other two are calculated using left or right half-face image and corresponding set of half-face eigenfaces. In Experiment 1 we show that performance of Eigenfaces recognition method using half-face images is worse than using full face images but still sufficient for identification purposes (drop from 92% to 84-85%).

If estimated head yaw angle φ_X of the subject being identified is bigger than a threshold Θ only a half of a synthesized face image is used for recognition. A feature vector is calculated using half-face eigenvectors and matched against a corresponding feature vectors of gallery images.

3 Experiments

In all experiments a training gallery consisting of $N = 2963$ normalized grey-level images from publicly available FERET database² and additional images of 3 subjects captured by us was used. All images represent faces in frontal position with neutral expression and captured under neutral lighting conditions. Each image was normalized using ground truth information about eyes position provided with FERET database. Original images were transformed, downsampled and cropped so that eyes have the same coordinates on all resultant images. Eyes position was chosen so that most of the background is removed but majority of the face area remains visible. At the end histogram equalization was applied. Resultant images have resolution 73×105 and examples can be seen on Fig. 3.

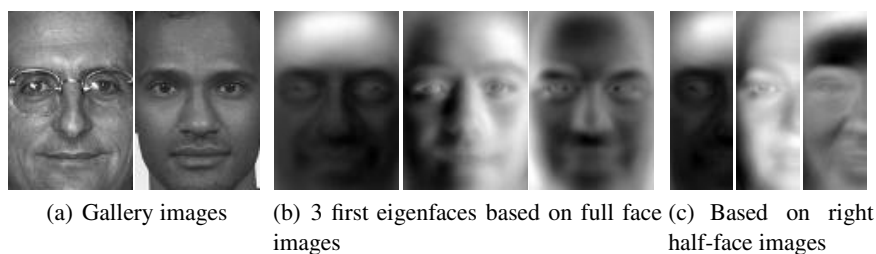


Fig. 3 Examples of gallery faces and eigenfaces

Experiment 1: Impact of using half-face images on Eigenfaces method

The aim of this experiment was to prove the feasibility of using half-face images for recognition. Due to intrinsic symmetry of a human face we expected that using only left or right half of human face image should not decrease recognition performance significantly.

Table 1 Impact of using half-face images on performance of Eigenfaces recognition method

Scenario	Number of correct identifications	Number of incorrect identifications	Percent of correct identifications
Full face image	1251	109	92%
Right half-face image	1142	208	84%
Left half-face image	1156	204	85%

A test set consisting of 1360 images, representing subjects from the training gallery, but acquired at a different time and having a slightly different expression, was chosen from FERET gallery. These images were normalized in the same way

² <http://www.itl.nist.gov/iad/humanid/feret/>

as images in a training gallery. Test set images were matched against gallery images using Eigenfaces method. 3 different scenarios were tested: using full face images, left half-face and right half-face images.

Results are summarized in Table 1. As can be seen, using only half-face images causes a decrease in recognition performance of Eigenfaces method by app. 9%. This means that using half-face image decreases discriminative ability of the face recognition system but when only a partial image of a face is available can be used for identification.

Experiment 2: Impact of head orientation variation on Eigenfaces method

Aim of this experiment was to verify an impact of a head rotation in a left-right direction (yaw angle) on a performance of Eigenfaces recognition method. A person sitting in front of a camera slowly rotated his head and a sequence of 33 images was captured, with yaw angle changing gradually from 0° (frontal view) to 64° (left profile view). Examples are depicted on Fig. 2(a). Distance between frontal view face (yaw angle = 0°) and each rotated face was calculated using Eigenfaces approach and results are depicted on Fig. 4. It can be seen that initially, when a yaw angle increases from 0° to 10° , the distance grows sharply and then steadily increases. Impact on recognition was verified in the following way: frontal view image (yaw angle = 0°) was added to our training gallery and each image of rotated face was matched against the gallery using Eigenfaces method. For small yaw angles, below 18° , images of rotated face were correctly matched with a frontal view image the same person. When yaw angle grew above 18° incorrect match was given.

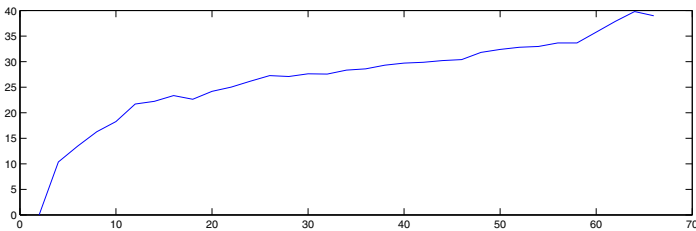


Fig. 4 Distance between *en face* and rotated face image as a function of yaw angle

Experiment 3: Impact of normalized en face view synthesis on Eigenfaces method

In this experiment we verified an impact of a normalized *en face* view synthesis step on a recognition system performance. A person sitting in front of a stereo camera set slowly rotated his head and a sequence of 33 image pairs was captured, with yaw angle changing gradually from 0° (frontal view) to 64° (left profile view). For each stereo image pair a shape of the head was recovered and a normalized *en face* view was rendered. Examples of the normalized view are shown on Fig. 2(b). Each synthesized image was matched against the training gallery containing a frontal view image of the subject face (acquired at 0° yaw angle) using Eigenfaces method.

Distance between frontal view image (captured at yaw angle = 0°) in the gallery and each synthesized *en face* view is depicted on Fig. 5. It can be seen that it grows slower compared to the distance to unnormalized faces in Experiment 2 so synthesized *en face* images resemble more similarity to the frontal view image in the gallery.

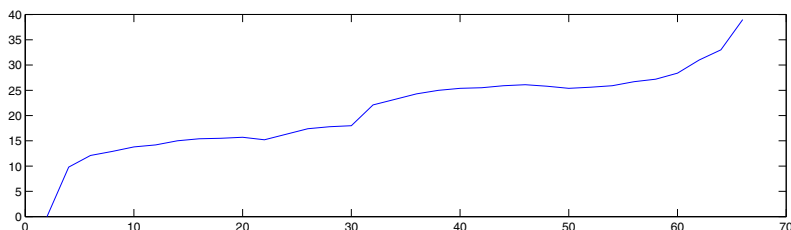


Fig. 5 Distance between gallery image (captured at yaw angle 0°) and synthesized *en face* image as a function of a yaw angle at which stereo pairs were captured

4 Conclusions and Future Work

The presented method can be used to increase resiliency of an image-based face recognition system to a head orientation although the improvement is moderate. For practical deployment the method must be further refined and a few of issues should be solved. Firstly an automatic head pose estimation subsystem should be implemented, as currently this step is done manually. In order to produce satisfactory results head pose estimation must be very accurate and published methods may prove not to be adequate. Additionally using stereo vision technique to reconstruct a shape of a human face from a single pair of stereo images has a limited accuracy. Noise and inaccuracies in shape reconstruction induce errors in normalized *en face* view synthesis which yields to incorrect recognition. As the head rotation angle increases impact of these errors becomes larger. Solutions to improve a quality of stereo reconstruction of a head shape should be investigated, e.g. by using sequence of stereo pairs from a camera set.

References

- [1] Blanz, W., Vetter, T.: Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(9), 1063–1074 (2003)
- [2] Fehn, C.: Depth-Image Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV. In: *Stereoscopic Displays and Virtual Reality Systems XI*, Proceedings of the SPIE, vol. 5291, pp. 93–104 (2004)
- [3] Hirschmuller, H.: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2005*, vol. 2, pp. 807–814 (2005)

- [4] Komorowski, J., Rokita, P.: A Method for Novel Face View Synthesis Using Stereo Vision. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010. LNCS, vol. 6375, pp. 49–56. Springer, Heidelberg (2010)
- [5] Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47, 7–42 (2002)
- [6] Schreer, O., Kauff, P., Sikora, T.: 3D Videocommunication, Algorithms, concepts, and real-time systems in human centered communication (2005)
- [7] Telea, A.: An image inpainting technique based on the fast marching method. *Journal of Graphics, GPU, and Game Tools* 9, 23–34 (2004)
- [8] Turk, M.A., Pentland, A.P.: Face Recognition Using Eigenfaces. In: *Proceedings of Computer Vision and Pattern Recognition, CVPR 1991*, pp. 586–591 (1991)
- [9] Voit, M., Nickel, K., Stiefelwagen, R.: Head Pose Estimation in Single- and Multi-view Environments – Results on the CLEAR07 Benchmarks. In: Stiefelwagen, R., Bowers, R., Fiscus, J.G. (eds.) RT 2007 and CLEAR 2007. LNCS, vol. 4625, pp. 307–316. Springer, Heidelberg (2008)
- [10] Zhao, W., Chellappa, R.: Image-based Face Recognition: Issues and Methods. In: Javidi, B. (ed.) *Image Recognition and Classification*, Marcel Dekker, New York Basel (2002)
- [11] Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.: Face Recognition: A Literature Survey. *ACM Computing Surveys* 35(4), 399–458 (2003)
- [12] Zhang, Z.: A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11), 1330–1334 (2000)

Subspace Algorithms for Face Verification

Maciej Smiatacz

Abstract. In real-life applications of face recognition the task of verification appears to be more important than classification. Usually we only have a limited collection of training images of some person and we want to decide if the acquired photograph is similar enough to them without using a separate set of negative samples. In this case it seems reasonable to use a subspace method based on a coordinate system constructed individually for the given class (person). This work presents two such methods: one based on SDF and the other inspired by Clafic algorithm. In the experimental section they are compared to the two-class SVM on the realistic data set taken from CMU-PIE database. The results confirm the advantages of subspace approach.

1 Introduction

In recent years automatic face recognition has turned into mature biometric technology. Some reports indicate that in particular cases computer systems recognize faces more accurately than humans [1]. Many commercial applications have appeared in this area, including the programs that provide visual login option even for low-end devices such as netbooks.

Although we have moved from experimental software to the mainstream technology, in terms of speed, accuracy and practical usefulness of automatic face recognition there is still much to improve. In real life many users get irritated by the fact that even small changes of pose or illumination usually cause identification failures. A functional face recognition system must be composed of several reliable modules performing image preprocessing, face detection, feature extraction, identification

Maciej Smiatacz

Faculty of Electronics, Telecommunications and Informatics,

Gdansk University of Technology, Narutowicza 11-12, 80-233 Gdansk, Poland

e-mail: slowhand@eti.pg.gda.pl

and other operations. In this paper we concentrate on the last stage of the process, i.e. the decision making algorithms.

Interestingly, although countless publications have focused on the problem of face *classification* in real life we have to cope with the problem of face *verification*, which is quite different. Namely, our task is not to choose from a set of known identities finding *the most similar* person; instead we have to decide, whether the acquired photograph is *similar enough* to the known images of a given person. In this case we take into account the two main parameters: false acceptance rate (FAR) that measures the percent of invalid inputs which are incorrectly accepted, and false rejection rate (FRR) measuring the percent of valid inputs which are incorrectly rejected. A good verification method should provide a tuning parameter allowing for easy adjustment of FAR to FRR ratio, depending on the type of application and user preferences.

In many cases the task of verification can be more difficult than simple classification. Although it is possible to treat the verification as a special case of two class discrimination (with one class representing “others”, i.e. containing the negative samples), sometimes we have to perform the identification using only a limited set of positive images.

In the following sections two face verification methods are proposed; both are based on the idea of *subspace* i.e. the coordinate system constructed specifically for a given class. In the experimental section their performance was compared to the results provided by a two class SVM algorithm.

2 Multi-axis SDF Method

The original SDF method was introduced in [2]. The core concept of this approach is the so-called Similarity Discriminant Function measuring the similarity of matrices \mathbf{Y}_i projected onto vector \mathbf{d} :

$$F(\mathbf{d}) = \frac{\mathbf{d}^T \left(\sum_{i=1}^m \sum_{j=1, i \neq j}^m \mathbf{Y}_i \mathbf{Y}_j^T \right) \mathbf{d}}{\mathbf{d}^T \left(\sum_{i=1}^m \mathbf{Y}_i \mathbf{Y}_i^T \right) \mathbf{d}} = \frac{\mathbf{d}^T \mathbf{B} \mathbf{d}}{\mathbf{d}^T \mathbf{C} \mathbf{d}} \quad (1)$$

where $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$ ($m \geq 2$) are the matrix patterns (e.g. typical 2D bitmap images) with dimensions $n_1 \times n_2$. The vector \mathbf{d} for which the function reaches the maximum value is the sought for pattern projection vector that we will denote as \mathbf{d}_{opt} . It can be proved that the higher the value of $F(\mathbf{d})$ the greater the concentration of images $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ projected onto \mathbf{d} .

Authors of the original paper suggested to calculate only one vector, \mathbf{d}_{opt} , for each class with the help of rank decomposition procedure [3] which was able to overcome the small sample size problem. On the other hand, they performed the secondary discriminant transformation based on multi-class Fisher criterion and finally applied the minimal distance classifier.

In [4] we proved that the secondary transformation only deteriorates the results. We also showed that by using more than one axis calculated from (1) the performance of the SDF procedure can be significantly improved. This led us to the formulation of the Multi-axis SDF classifier which is a true subspace method as the individual coordinate system is calculated for each class. In the case of verification problem we have only one class of images (the training set contains images of only one person). The subspace can be constructed in the following way.

Algorithm 1 - construction of the Multi-axis SDF subspace

1. Calculate matrices $\mathbf{B} = \sum_{i=1}^m \sum_{j=1, i \neq j}^m \mathbf{Y}_i \mathbf{Y}_j^T$ and $\mathbf{C} = \sum_{i=1}^m \mathbf{Y}_i \mathbf{Y}_i^T$ using m training images $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$.
2. On the basis of the SDF function (1) calculate the projection axes $\mathbf{d}_{opt(a)}$ ($a = 1, \dots, n_a$) using the rank decomposition approach. To this end the matrix \mathbf{P} containing eigenvectors of \mathbf{C} must be created, then the following transformation of matrix \mathbf{B} is performed:

$$\bar{\mathbf{B}} = (\bar{\mathbf{A}}\mathbf{P})\mathbf{B}(\bar{\mathbf{A}}\mathbf{P})^T \quad (2)$$

where matrix $\bar{\mathbf{A}} = \text{diag}(\bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_N)$

$$\bar{\lambda}_k = \begin{cases} 1/\sqrt{\lambda_k} & , \lambda_k > 0 \\ 1 & , \lambda_k = 0, \end{cases} \quad (3)$$

and λ_k denotes the k -th eigenvalue of \mathbf{C} . The optimal projection axes $\mathbf{d}_{opt(a)}$ can be calculated from the subsequent eigenvectors of $\bar{\mathbf{B}}$ [3].

3. For each image \mathbf{Y}_i ($i = 1, \dots, m$) calculate the feature vectors $\mathbf{y}_i^{(a)}$:

$$\mathbf{y}_i^{(a)} = \mathbf{Y}_i^T \mathbf{d}_{opt(a)}, \quad a = 1, \dots, n_a \quad (4)$$

4. Compute the mean vectors along all axes:

$$\boldsymbol{\mu}^{(a)} = \frac{1}{m} \sum_{i=1}^m \mathbf{y}_i^{(a)}, \quad a = 1, \dots, n_a \quad (5)$$

Having created the subspace, we can perform the verification of \mathbf{Y} as follows.

Algorithm 2 - Multi-axis SDF verification method

1. Calculate the feature vectors $\mathbf{y}^{(a)}$ corresponding to the subsequent axes $\mathbf{d}_{opt(a)}$:

$$\mathbf{y}^{(a)} = \mathbf{Y}^T \mathbf{d}_{opt(a)}, \quad a = 1, \dots, n_a \quad (6)$$

2. Determine the level of similarity of the image \mathbf{Y} to the class:

$$D(\mathbf{Y}) = \sum_{a=1}^{n_a} w^{(a)} D^{(a)}(\mathbf{Y}) \quad (7)$$

where

$$D^{(a)}(\mathbf{Y}) = \sqrt{(\mathbf{y}^{(a)} - \boldsymbol{\mu}^{(a)})^T (\mathbf{y}^{(a)} - \boldsymbol{\mu}^{(a)})} \quad (8)$$

$w^{(a)}$ - a weight of the axis $\mathbf{d}_{opt(a)}$ that can be determined using the eigenvalues λ_k of matrix \mathbf{C} :

$$w^{(a)} = \lambda_a \left/ \sum_{k=1}^{n_a} \lambda_k \right. \quad (9)$$

3. Classify the image as positive if the following condition is satisfied:

$$D(\mathbf{Y}) < \alpha \cdot \max_{i=1..m} D(\mathbf{Y}_i) \quad (10)$$

By changing the value of multiplier α we can tune the FAR/FRR ratio of the method.

3 Two-Stage Clafic Algorithm

The original Clafic method was proposed by Watanabe in [5]. This classification algorithm discards the between-class dependencies and the coordinate system is optimized individually for each class c_l , $l = 1, 2, \dots, L$. The correlation matrix $\mathbf{C}^{(l)}$ of class c_l is defined as follows:

$$\mathbf{C}^{(l)} = \frac{1}{m_l} \sum_{i=1}^{m_l} \mathbf{y}_i \mathbf{y}_i^T \quad (11)$$

where m_l denotes the number of vector patterns \mathbf{y}_i in the training set of class c_l ; the patterns are normalized without the centralization, i.e. the subtraction of mean vectors of the classes is not performed. Watanabe points out that the mean vectors are not used in any way during the construction of the matrix (11).

If the matrix that transforms the feature space is composed from the eigenvectors of matrix $\mathbf{C}^{(l)}$ sorted by their descending eigenvalues $\lambda_k^{(l)}$, then the base vectors with the lowest indices represent the most characteristic properties of class c_l . In other words, the information about class c_l is concentrated in the first (the most important) coordinates of the new system. From this fact Watanabe draws the conclusion that the entropy of the components of feature vectors describing the objects from class c_l reaches minimum in the system [6] and so we can expect that in the case of objects that do not belong to class c_l the information will be more uniformly distributed among all the components of the feature vector, and the entropy, as a measure of the information scatter, will take larger values. Thus, the entropy can be used for classification purposes.

The above considerations form the basis of the classification procedure. Watanabe called this variant of the Clafic method the dispersion algorithm [5]. It can be adapted to the problem of verification in the following way.

Algorithm 3 - The Clafic verification method, training

1. Perform the normalization of vectors \mathbf{y}_i from the training set and create the correlation matrix (11).
2. Using the correlation matrix \mathbf{C} find the optimal coordinate system by determining the eigenvectors $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N$, ordered by their eigenvalues, $\lambda_1 > \lambda_2 > \dots > \lambda_N$.
3. Transform the training set to the new space:

$$\mathbf{v}_i = \mathbf{G}^T \mathbf{y}_i, \quad \mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N] \quad (12)$$

4. For every \mathbf{v}_i ($i = 1, \dots, m$) calculate the entropy of its components:

$$H(\mathbf{v}_i) = - \sum_{k=1}^N v_{ik}^2 \log v_{ik}^2 \quad (13)$$

5. Compute the maximum entropy in the training set:

$$H_t = \max_{i=1..m} H(\mathbf{v}_i) \quad (14)$$

The value of (14) will be used in the testing algorithm.

Algorithm 4 - The Clafic verification method, testing

1. Normalize the unknown object \mathbf{y} and convert it to \mathbf{v} , according to (12), performing transformation to the space spanned by eigenvectors $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N$.
2. Calculate the entropy $H(\mathbf{v})$ of the components of \mathbf{v} (13).
3. Classify the image as positive if the following condition is satisfied:

$$H(\mathbf{v}_i) \leq H_t + \alpha \quad (15)$$

where α is a tuning parameter.

If we want to treat the input images \mathbf{Y}_i directly by using each pixel intensity as a separate feature the above algorithms become very slow. This is because each bitmap \mathbf{Y}_i must be converted to a large \mathbf{y}_i vector by concatenating the subsequent rows of \mathbf{Y}_i . In our experiments we used 50×50 face images so each \mathbf{y}_i vector contained 2500 elements. Even though the eigenvectors can be efficiently calculated by using the scrambled version of matrix (11), as described, for example, in [7], the conversions and projections of \mathbf{y}_i vectors take a lot of time.

Therefore we decided to speed up the process utilizing the idea of creating the correlation matrix (11) directly from \mathbf{Y}_i matrices, without converting them to vectors

first (see, for example [8] or [9]). This way the following Two-stage Clafic algorithm was formulated.

Algorithm 5 - Two-stage Clafic verification method, training

1. Create the correlation matrix (no special normalization of training patterns \mathbf{Y}_i is performed):

$$\mathbf{C}_M = \frac{1}{m} \sum_{i=1}^m \mathbf{Y}_i \mathbf{Y}_i^T \quad (16)$$

2. Calculate K eigenvectors \mathbf{g}_{Mk} , $k = 1, 2, \dots, K$.
3. Transform the training set to the new space

$$\mathbf{y}_i^{(k)} = \mathbf{Y}_i^T \mathbf{g}_{Mk}, \quad k = 1, 2, \dots, K, \quad (17)$$

4. Create vectors \mathbf{y}_i , each of which is obtained by the concatenation of the set of vectors $\{\mathbf{y}_i^{(k)}\}$.
5. Create the classic Watanabe's classifier, treating vectors \mathbf{y}_i as the source data for Clafic method (see Algorithm 3). This means that we have to calculate the transformation matrix \mathbf{G} (12).

As we can see, K is a parameter controlling the information reduction at the first stage. As a result of the transformations performed at points 3 and 4 the \mathbf{y}_i vectors will contain $K \cdot x$ elements where x is the number of columns in \mathbf{Y}_i .

4 Experiments

The algorithms described in the previous sections were implemented in the newest version of *SimPaRS*, our experimental Windows application written in C++. We also used other C++ programs based on OpenCV 2.1 library that contained the implementation of Support Vector Machine and AdaBoost cascaded face detector.

The tests were performed on the data coming from CMU-PIE database [10]. This set contains realistic images taken in diverse illumination conditions from several viewpoints. Different face expressions are visible, some persons wear glasses on selected pictures. The photographs showing profile views or taken under extremely poor lighting were removed from the set.

The AdaBoost algorithm based on Haar-like features [11] was used to crop the faces from the background. The method, however, failed to detect faces on some of the pictures. Finally, our database contained 13153 face images of 68 persons. They were scaled down to the resolution of 50×50 pixels, the histogram normalization was also performed.

The images taken from two almost frontal viewpoints (cameras no 7 and 27) during the *expression* session (images of the subjects with a neutral expression, smiling, and blinking) were selected as the training set. The number of training images per person ranged from 4 to 8. The system was trained with this small amount of images



Fig. 1 Selected training samples (upper row) and testing images (lower row) of one person

of a given person and then the whole database containing about 200 “positive” and 13000 “negative” examples was used during the test. This procedure was repeated 68 times (for all the persons). Sample images from the database are shown in Fig. 1.

The above verification task was difficult due to large variations in illumination, pose and expression. Popular face identification software, such as the visual login applications shipped with many notebooks today, would fail in most of the positive cases, resulting in the very high false rejection rate.

The two-class Support Vector Machine with polynomial kernel was used as a reference solution. It is important to notice, however, that this way the subspace algorithms were confronted with a competitor that was impossible to beat. First of all the SVM was trained within “one against the others” scenario, so the algorithm was able to collect the knowledge about the “intruders” at the training stage; moreover, selected images of *all* the intruders were presented during training, although it was only a small subset of the testing database. In other words, the SVM used much more information than the subspace methods; additionally, it was based on nonlinear kernel. Therefore, it was not surprising that the SVM performed better; the subspace methods, however, require only a small training set of positive samples and their FAR/FRR ratio can be easily tuned by changing the appropriate parameters that are not present in SVM.

Table 1 Average false rejection and false acceptance rates, and the time needed to perform the whole test (68 classes \times 13153 images) for different verification methods

	SVM	Multi-axis	SDF Clafic	Two-stage Clafic
FRR (%)	58.47	75.03	75.58	76.77
FAR (%)	0.06	0.04	0.04	0.05
testing time (s)	473	690	4980	1020

The results of experiments are summarized in Table 1. The parameters of the methods were set to achieve the very low FAR of SVM (about 5 out of 10000 intrusion trials were successful). Under this conditions all of the subspace algorithms performed equally, providing the FRR of about 75%, which was high but not so dramatically worse than the result reached by SVM. The original Clafic method is particularly slow because it has to operate on large feature vectors. By introducing the additional stage involving the information selection the verification time (for the whole experiment) was reduced by 5 times from 4980s to 1020s.

5 Conclusion

In this paper we proposed the Multi-axis SDF and the Two-stage Clafic verification methods. They are both based on subspaces and require only a small set of positive training samples to build the identification system. Additionally, their performance can be adjusted so when the user demands a more comfortable solution at the lower security level, it can be achieved with the simple threshold manipulation.

Our experiments show that the proposed subspace methods can be used in a face verification system. It is important to notice, however, that we tried to use difficult to analyze, real-life data, with very limited preprocessing. A specialized biometric application would employ a better segmentation algorithm, a module minimizing the influence of illumination changes, and a sophisticated feature extraction algorithm. Under these conditions the performance of the presented methods would be significantly better.

Acknowledgement. This work has been supported by the Polish Ministry of Science and Higher Education as the research project No. N N516 36793.

References

- [1] Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: FRVT 2006, and ICE 2006, Large-Scale Results. NISTIR 7408 National Institute of Standards and Technology, Gaithersburg (2007)
- [2] Cheng, Y.-Q., Liu, K., Yang, J.-Y.: A Novel Feature Extraction Method for Image Recognition Based on Similar Discriminant Function (SDF). *Pattern Recognition* 26(1), 115–125 (1993)
- [3] Cheng, Y.-Q., Zhuang, Y.-M., Yang, J.-Y.: Optimal Fisher Discriminant Analysis Using the Rank Decomposition. *Pattern Recognition* 25(1), 101–111 (1992)
- [4] Smiatacz, M., Malina, W.: New variants of the SDF classifier. In: Kurzynski, M., Wozniak, M. (eds.) *Computer Recognition Systems 3*. AISC, vol. 57, pp. 205–212. Springer, Heidelberg (2009)
- [5] Watanabe, S., Lambert, P.F., Kulikowski, C.A., Buxton, J.L., Walker, R.: Evaluation and Selection of Variables in Pattern Recognition. *Computer and Information Sciences II*, 91–122 (1967)

- [6] Watanabe, S.: Karhunen-Loeve Expansion and Factor Analysis. In: Trans. Fourth Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes, pp. 635–660 (1965)
- [7] Turk, M., Pentland, A.: Eigenfaces for Recognition. *J. Cognitive Neuroscience* 3(1), 71–86 (1990)
- [8] Smiatacz, M., Malina, W.: Modifying the Input Data Structure for Fisher Classifier. In: Proc. Second Conf. on Computer Recognition Systems (KOSYR 2001), pp. 363–367 (2001)
- [9] Yang, J., Zhang, D., Frangi, A.F., Yang, J.Y.: Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence* 26(1), 131–137 (2004)
- [10] Sim, T., Baker, S., Bsat, M.: The CMU Pose, Illumination, and Expression (PIE) Database. In: Proc. 5th Int. Conf. on Automatic Face and Gesture Recognition (2002)
- [11] Viola, P., Jones, M.J.: Robust Real-Time Face Detection. *Int. J. Comp. Vision* 57(2), 137–154 (2004)

Fingerprint Orientation Field Enhancement

Lukasz Wieclaw

Abstract. This paper presents a new method to enhance the fingerprint orientation image. Orientation, as a global feature of fingerprint, is very important to image preprocessing methods used in automatic fingerprint identification systems (AFIS). The most popular, gradient-based method is very sensitive to noise (image quality). Proposed algorithm is an application of gradient-based method combined with more resistant to noise pixel-alignment-based method. Experimental results show that the proposed method is robust to noise and still maintaining accurate values in high-curvature areas.

1 Introduction

Among various biometric techniques, automatic fingerprint based systems are regarded as most popular and reliable for automatic personal identification. With the increasing attention on automatic identity verification, fingerprint recognition systems have become a popular research topic over the last decades. However, there still exist critical research issues such as the low performance with poor quality images [1]. This problem is solved by image enhancement processing. Nevertheless, this step requires information about local orientation of fingerprint pattern. Fingerprint pattern consist of ridges and valleys on the surface of a fingertip. On a typical fingerprint image, the ridges are black and the valleys are white. Generally, fingerprints contain two kinds of features: local features like minutia or singular points (core and delta) and global features, such as a ridge pattern orientation and frequency. The local ridge structures are commonly used to identify or verify person

Lukasz Wieclaw

University of Silesia, Institute of Computer Science, 41-200 Sosnowiec,
Bedzińska 39, Poland

e-mail: lukasz.wieclaw@us.edu.pl

identity by matching two sets of minutiae. As a global feature, orientation field describes local orientation of the ridgevalley structure in each point of fingerprint image (Fig 1).

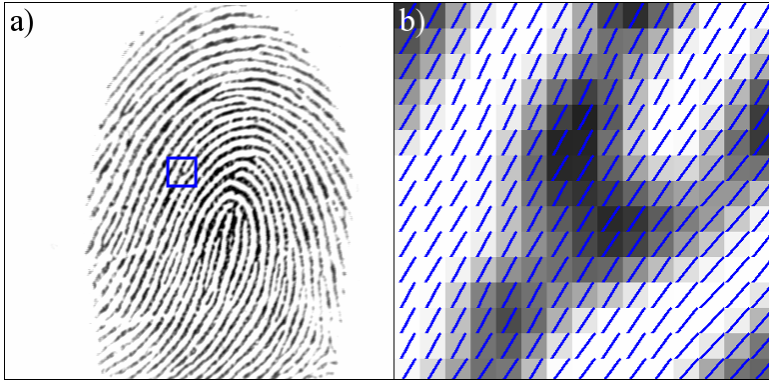


Fig. 1 a) Original fingerprint image, b) magnified area with marked dominant orientation of each pixel

The values of orientation angles of fingerprint pattern have a critical impact on almost all subsequent processes in automatic fingerprint recognition systems. Orientation field has been widely used for fingerprint image enhancement [2][3], singular points detection [4][5][6] and classification [7][8][9][10]. However, there are many low quality fingerprint images caused by poor skin condition (scars, wet or dry fingers), noisy acquisition devices or bad imprint techniques. Therefore, orientation values cannot always be obtained accurately. Due to the fact that the orientation estimation is usually the first stage of fingerprint image analysis, orientation field enhancement has become a necessary and common step before image enhancement and feature extraction in the AFIS.

Generally, there are two categories of methods to compute the orientation field (also called directional field or directional image): pixel-alignment based [5][11] and gradient-based methods [2][12]. Typically, the pixel-alignment based methods compute the differentiation (fluctuation) of neighboring pixels values in a fixed number of reference orientations (Fig. 2b). The most popular gradient-based method is presented by M. Kass and A. Witkin [12] least squares contour alignment method. The most important advantage of this algorithm is the fact, that the obtained values are continuous. The most advanced orientation field enhancement method is the model-based method [13][14], which relies on the global regularity of orientation values around the singular points. However, forensic images of latent fingerprints not always contain those singular points. Therefore, method proposed in this paper is filtering-based, that combines very accurate gradient-based method with the more resistant to noise pixel-alignment-based method.

2 Normalization of Fingerprint Image

In the presented method, before the gradient based calculation of orientation field, the image normalization and ridge pattern enhancement step is performed. Normalization is performed to remove the effects of sensor noise or inconsistency in ridge pattern contrast (Fig 2).

Let $I(x, y)$ denote the gray value at pixel (x, y) , $min_v(x, y)$ and $max_v(x, y)$ are the minimal and maximal values in $W \times W$ window centered at (x, y) , and $N(x, y)$ the normalized gray-level value. For all the pixels in $W \times W$ window the normalized pixel value is defined as:

$$N(x, y) = \frac{255(I(x, y) - min_v(x, y))}{max_v(x, y) - min_v(x, y)} \quad (1)$$

where:

$$min_v(x, y) = \min\{I(i, j), i = x - W, \dots, x, \dots, x + W, j = y - W, \dots, y, \dots, y + W\} \quad (2)$$

$$max_v(x, y) = \max\{I(i, j), i = x - W, \dots, x, \dots, x + W, j = y - W, \dots, y, \dots, y + W\} \quad (3)$$

The size of the $W \times W$ window must be larger than mean distance between fingerprint ridges. Experimental results showed that the $W = 8$ is optimal for the 512dpi fingerprint images.

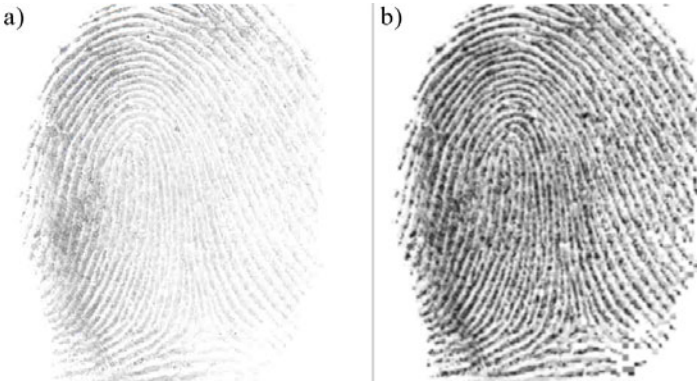


Fig. 2 a) Original low contrast fingerprint image, b) Normalized fingerprint image.

3 Ridge Pattern Enhancement

Filtering-based orientation estimation methods only operate at the local region and thus they cannot solve large noisy (wet or dry fingerprints) or missing patches (scars) in the image. Therefore, in presented method, the ridge pattern enhancement is performed by oriented anisotropic Gaussian filter, before final orientation computation.

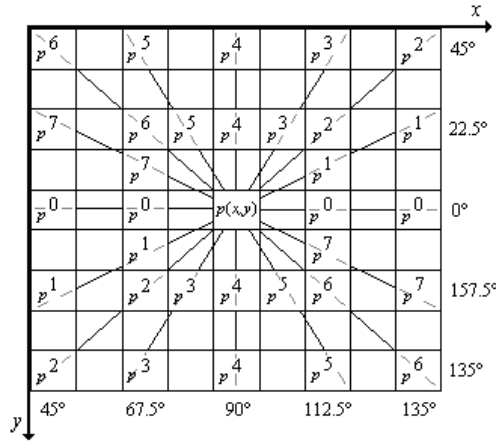


Fig. 3 The 9×9 mask to compute the differentiation of pixel values.

The orientation of Gaussian filter is obtained by robust to noise pixel-alignment method. The orientation is computed by a mask of eight oriented differentiations of pixel values. The structure of this 9×9 mask (Fig 3) was presented in [15] and was specially developed to obtain dominant direction of fingerprint ridge valley structure. Estimated values were limited to fixed number of eight discrete values, therefore this modification uses this mask only to compute eight differentiations of pixel values.

In the first stage, the mean values of five pixels in eight directions are computed:

$$S_i(x,y) = \frac{\sum_{j=1}^5 p_j^i}{5} \quad (4)$$

where:

$S_i(x,y)$ - is the means of pixel values in eight directions,

p_j^i - is the pixel values in one of i directions from the normalized image,

i - is the discrete direction value (0, ..., 7), respectively from 0° to $157,5^\circ$, with $22,5^\circ$ step.

In the next stage, the differentiation (fluctuation) of neighboring pixels values is computed, in each direction, by:

$$Df_i(x,y) = \sum_{j=1}^5 |S_i(x,y) - p_j^i|, \quad i = 0, 1, \dots, 7 \quad (5)$$

The orientation of lowest fluctuation of gray values is expected to be the reference orientation of the central pixel.

Due to the presence of some unreliable elements, resulting from heavy noise, corrupted ridge and furrow structures, minutiae and low gray value contrast, the estimated differentiation values may not always be correct. The orientation smoothing stage is expected to reduce the noise and compute a reliable values.

The averaging is computed for each pixel in 13×13 window, separately in each direction values, respectively:

$$Ad_i(x, y) = \sum_{u=x-\frac{W}{2}}^{x+\frac{W}{2}} \sum_{v=y-\frac{W}{2}}^{y+\frac{W}{2}} Df_i(u, v), \quad i = 0, 1, \dots, 7 \quad (6)$$

where $W = 13$.

The orientation of the smallest value, from all eight oriented averaged differentiation values, is expected to be the most closest to the dominant orientation of that pixel:

$$\alpha(x, y) = i_{min}(x, y)22,5^\circ \quad (7)$$

where:

$$i_{min}(x, y) = \arg(\min\{Ad_i(x, y), \quad i = 0, 1, \dots, 7\}) \quad (8)$$

Finally, the ridge pattern enhancement is performed by anisotropic Gaussian filtering:

$$En(x, y) = \frac{\sum_{u=-W}^W \sum_{v=-W}^W N(u+x, v+y)G_a(u, v)}{\sum_{u=-W}^W \sum_{v=-W}^W G_a(u, v)} \quad (9)$$

where $W = 4$ and anisotropic Gaussian kernel is given by:

$$G_a(u, v) = \exp\left(\frac{(u\alpha C)^2 + (v\alpha S)^2}{2\sigma_i^2} + \frac{(-u\alpha S)^2 + (v\alpha C)^2}{2\sigma_j^2}\right), \quad (10)$$

$$\alpha C = \cos \alpha(x, y),$$

$$\alpha S = \sin \alpha(x, y),$$

where $\sigma_i = 10$, $\sigma_j = 90$ and their ratio determines the Gaussian kernel flattening deformation.

The ridge pattern enhancement is expected to reduce the noise and influence of wet or dry patterns and scars on fingerprint images (Fig 4).

4 Orientation Field Estimation

Tasks, like tracing finger furrows lines or computing the ridge frequency, require continuous values of orientation field. Therefore the most frequently adopted method [1],[2] is the least mean square method [12]. In the presented paper, the modified least mean square method is presented.

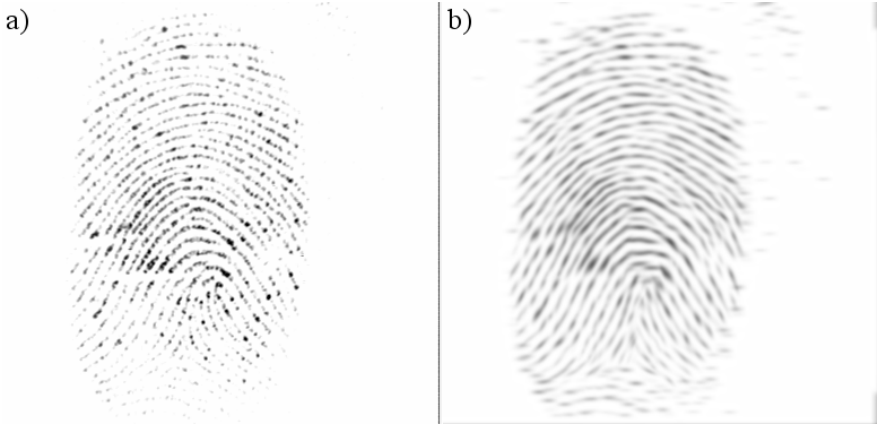


Fig. 4 a) Original low fingerprint image, b) Fingerprint image after normalization and ridge enhancement.

Since the gradients are the orientations at the pixel scale, the orientation of the ridge is orthogonal to the average phase angle of pixels value changes, indicated by gradients. The main steps of the modified least mean square algorithm are as follows:

1. Compute the gradients $\partial_x(x, y)$ and $\partial_y(x, y)$ at each pixel of the fingerprint image $En(x, y)$. Depending on the computational requirement, the gradient operator may vary from the simple Sobel operator to the more complex Marr-Hildreth operator.
2. If gradient values are the same $\partial_x(x, y) = \partial_y(x, y)$ then add randomly ± 1 to one of the gradients. If one of gradients values is equal to 0 (for example $\partial_x(x, y) = 0$) then also randomly ± 1 .
3. Estimate the local orientation in $W \times W$ blocks, centered at pixel (x, y) using the following equations:

$$\mathcal{V}x(x, y) = \sum_{u=x-\frac{w}{2}}^{x+\frac{w}{2}} \sum_{v=y-\frac{w}{2}}^{y+\frac{w}{2}} 2\partial_x(u, v)\partial_y(u, v) \quad (11)$$

$$\mathcal{V}y(x, y) = \sum_{u=x-\frac{w}{2}}^{x+\frac{w}{2}} \sum_{v=y-\frac{w}{2}}^{y+\frac{w}{2}} (\partial_x^2(u, v) - \partial_y^2(u, v)) \quad (12)$$

$$\theta(x, y) = \frac{1}{2} \tan^{-1} \left(\frac{\mathcal{V}x(x, y)}{\mathcal{V}y(x, y)} \right) \quad (13)$$

5 Experimental Results

The proposed algorithm had been applied on the fingerprint images database, which contains 960 fingerprint images (in 500dpi resolution) from 120 fingers, with 8 images from each finger. Fingerprint images in the database were selected from NIST Special Database, FVC database samples, and captured with a live scanner.

In order to obtain the performance, the proposed method was compared with the most popular gradient-based enhancement method proposed in [2].

The mean difference between orientation estimated by the automatic method and orientation estimated by a police expert, was measured. The results summarized in (Fig 5) show that the proposed orientation field enhancement method has a 12% lower error level in comparison method presented in [2].

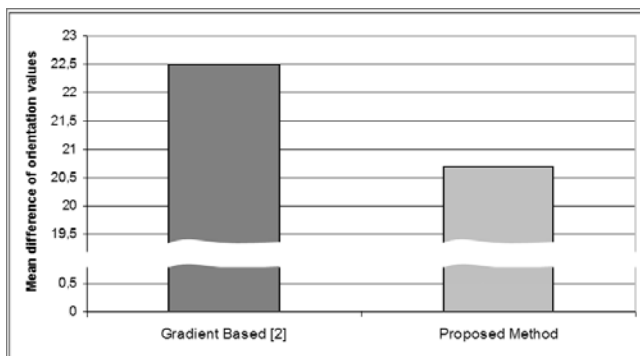


Fig. 5 Mean difference between orientation estimated by automatic methods and orientation estimated by police expert.

6 Conclusion

In this paper a new method for the fingerprint orientation field estimation has been presented.

It can be concluded from evaluated experiments, that by using the proposed algorithm, the estimated results are accurate and robust to noise.

Further work will consider the modification of differentiation averaging step in pixel-alignment ridge pattern enhancement.

Acknowledgment. This paper was supported by the grant no. N N519 574038 from Ministry of Science and Higher Education, Poland.

References

- [1] Jain, A.K., Bolle, R., Pankanti, S.: Biometrics: Personal Identification in Networked Society. Kluwer, New York (1999)
- [2] Hong, L., Jain, A.K., Wan, Y.: Fingerprint Image Enhancement: Algorithm and Performance Evaluation. IEEE Trans. on Pattern Analysis and Machine Intelligence 20(8), 777–789 (1998)

- [3] Chikkerur, S., Cartwright, A.N., Govindaraju, V.: Fingerprint enhancement using STFT analysis. *Pattern Recogn.* 40(1), 198–211 (2007)
- [4] Liu, M., Jiang, X., Kot, A.C.: Fingerprint reference-point detection. *EURASIP J. Appl. Signal Process.*, 498–509 (2005)
- [5] Bazen, A.M., Gerez, S.H.: Systematic Methods for the Computation of the Directional Fields and Singular Points of Fingerprints. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(7), 905–919 (2002)
- [6] Wrobel, K., Doroz, R.: New Method For Finding a Reference Point in Fingerprint Images With the Use Of The IPAN99 Algorithm. *Journal of Medical Informatics & Technologies* 13, 59–64 (2009)
- [7] Hong, L., Jain, A.K., Prabhakar, S.: A Multichannel Approach to Fingerprint Classification. *IEEE Trans. Pattern Anal. Mach. Intell.* 21(4), 348–359 (1999)
- [8] Costa, S.M., Fernandez, F.J., Oliveira, J.M.: A New Paradigm on Fingerprint Classification using Directional Image. In: *SIBGRAPI*, p. 405 (2002)
- [9] Jain, A.K., Karu, K.: Fingerprint classification. *Pattern Recognition* 29(3), 38–44 (1996)
- [10] Halici, U., Ongun, G.: Fingerprint classification through self-organizing feature maps modified to treat uncertainties. *Proc. of the IEEE* 84(10), 1497–1512 (1996)
- [11] Hong, L., Jain, A.K., Pankanti, S., Prabhakar, S.: Filterbank-based fingerprint matching. *IEEE Trans. Image Processing* 9(5), 846–859 (2000)
- [12] Kass, M., Witkin, A.: Analyzing Orientated Pattern. *Computer Vision, Graphics and Image Processing* 37, 362–397 (1987)
- [13] Gu, J., Zhou, J.: Modeling orientation fields of fingerprints with rational complex functions. *Pattern Recognition* 37(2), 389–391 (2004)
- [14] Li, J., Wang, H., Yau, W.: Nonlinear Phase Portrait Modeling of Fingerprint Orientation. In: *IEEE Proc. Control, Automation, Robotics and Vision Conf.*, vol. 2, pp. 1262–1267 (2004)
- [15] Porwik, P., Wieclaw, L.: A new approach to reference point location in fingerprint recognition. *IEICE Int. Journal Electronics Express* 1(18), 575–581 (2004)

Lip Print Recognition Based on Mean Differences Similarity Measure

Lukasz Smacki and Krzysztof Wrobel

Abstract. This paper presents a new method for comparing and classifying lip print images. In the proposed method a lip print image is first binarized and then subjected to the Hough transform. As a result a collection of line segments approximating the lip print pattern is obtained. Each segment is described by its length, angle and midpoint coordinates. Lip prints are compared using the mean differences similarity measure. Presented studies tested the impact of different weights applied to segment's characteristic features on lip print recognition results. After further improvements the presented method can be used in criminal identification systems.

1 Introduction

Lip prints are impressions of human lips left on objects such as drinking glasses, cigarettes, drink containers, etc. French criminologist, Edmond Locard, first recommended the use of lip prints for criminal identification in 1932 [9]. Study of human lips, as a means of personal identification, was started in 1970s by two Japanese scientists Yasuo Tsachihasi and Kazuo Suzuki [13, 16].

Forensic investigation technique that deals with identification of humans based on lip prints is called cheiloscopy [12]. This form of identification is slowly accepted and introduced to practice all over the world. In many countries, including Poland, lip traces are successfully used as evidence in lawsuits [8, 11].

Surface of lips is covered by numerous depressions (furrows), which reflect as a line pattern. They are arranged in various shapes and drawings, which are individual for each person. There are no two people with identical lip furrows arrangement [2]. Personal identification based on lip traces is possible due to the fact that lip pattern,

Lukasz Smacki · Krzysztof Wrobel

Institute of Computer Science, University of Silesia, ul. Bedzinska 39, 41-200 Sosnowiec

e-mail: lukasz.smacki,krzysztof.wrobel@us.edu.pl

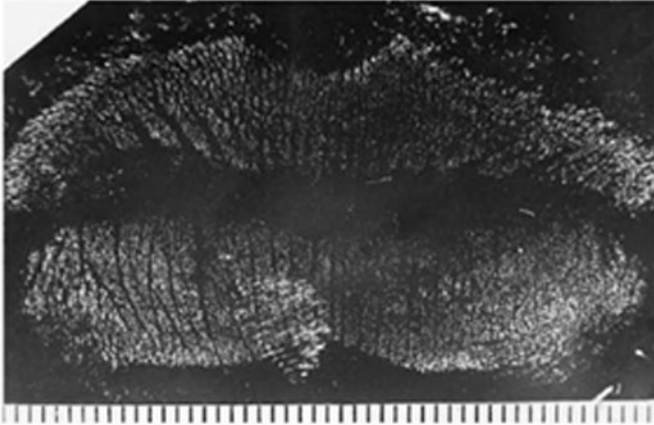


Fig. 1 Sample lip print recovered from a crime scene.

similar to fingerprint pattern, has three properties: permanence, indestructibility and uniqueness [2, 8, 16]. Figure 1 presents an example of a lip print recovered from a crime scene.

2 Image Processing

Lip print image processing deals with extracting a set of characteristics of a lip print. The set of characteristics is a collection of segments, which approximates a lip pattern present on the analyzed lip print. Image processing procedure used in the proposed method consists of two main stages (presented in detail in previous articles [13, 14]):

- Pre-processing, where lip pattern is extracted from lip print image and converted to black color,
- Feature extraction, where lip pattern is approximated to a corresponding collection of segments.

Background detection is the first stage of the image pre-processing. The aim of this procedure is to separate lip print area from the background. This way it is possible to use different types of image processing for these two areas of the image. Each pixel of an original grey scale lip print image I_{ORG} (Figure 2a) is checked whether it belongs to background or lip print area. Pixel $I_{ORG}(x,y)$ belongs to the image background, if the average brightness of pixels in its surroundings $d(x,y)$ is greater than the background detection threshold γ :

$$I_{BG}(x,y) = \begin{cases} 0 & \text{if } d(x,y) > \gamma \\ I_{ORG}(x,y) & \text{otherwise} \end{cases} \quad (1)$$

where:

$$x = 1, 2, \dots, w,$$

$$y = 1, 2, \dots, h,$$

w - width of image I_{ORG} ,

h - height of image I_{ORG} .

Parameters $d(x, y)$ and γ are defined by the following formulas:

$$\gamma = 255 - 0.75 \left(255 - \frac{\sum_{y=1}^h \sum_{x=1}^w I_{ORG}(x, y)}{wh} \right) \quad (2)$$

where:

$$x = 1, 2, \dots, w,$$

$$y = 1, 2, \dots, h.$$

$$d(x, y) = \frac{\sum_{j=y-1}^{y+1} \sum_{i=x-3}^{x+3} I_{ORG}(i, j)}{21} \quad (3)$$

where:

$$x = 4, 5, \dots, w - 3,$$

$$y = 2, 3, \dots, h - 1.$$

The second stage of the image pre-processing is binarization. The main goal of the binarization procedure is to remove image artefacts and emphasize lip pattern for further processing. Binarization is carried out for every pixel of image I_{BG} . If a pixel is part of the background, it takes the value of 0 (white). Otherwise its value is determined as follows:

$$I_{BIN}(x, y) = \begin{cases} 1 & \text{if } I_{BG}(x, y) > b(x, y) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where:

$$x = 1, 2, \dots, w,$$

$$y = 1, 2, \dots, h.$$

Parameter $b(x, y)$ is defined by the following formula:

$$b(x, y) = 1.1 \left(\frac{\sum_{j=y-4}^{y+4} \sum_{i=x-4}^{x+4} I_{BG}(i, j)}{81} \right) \quad (5)$$

where:

$$x = 5, 6, \dots, w - 4,$$

$$y = 5, 6, \dots, h - 4.$$

The final result of binarization is presented in Figure 2b. After the pre-processing, image feature extraction is conducted. First stage of the feature extraction is Hough

transform [4, 5, 6]. This method is used to detect straight lines in the lip imprint image. In the first step pixels of image I_{BIN} belonging to lip pattern area (marked black in Figure 2b) are transferred to polar coordinate system. On the basis of I_{BIN} coordinates, length of vector r for each value of angle θ is determined:

$$r = x \cos \theta + y \sin \theta, \quad \theta \in [0, \pi). \quad (6)$$

Obtained pair of polar coordinates allocates an address of cell in a so-called accumulator array. Value in the allocated cell is then incremented. This way pixels belonging to lip print pattern are reproduced in the accumulator array as curves.

In the second step, intersections of curves in the accumulator array are identified. Coordinates of the intersection points determine the parameters of straight lines existing in the lip pattern. Analysis is conducted to all cells in the accumulator array. Coordinates of cells that form the local maxima define the (r, θ) parameters used to determine straight lines equations according to formula 6.

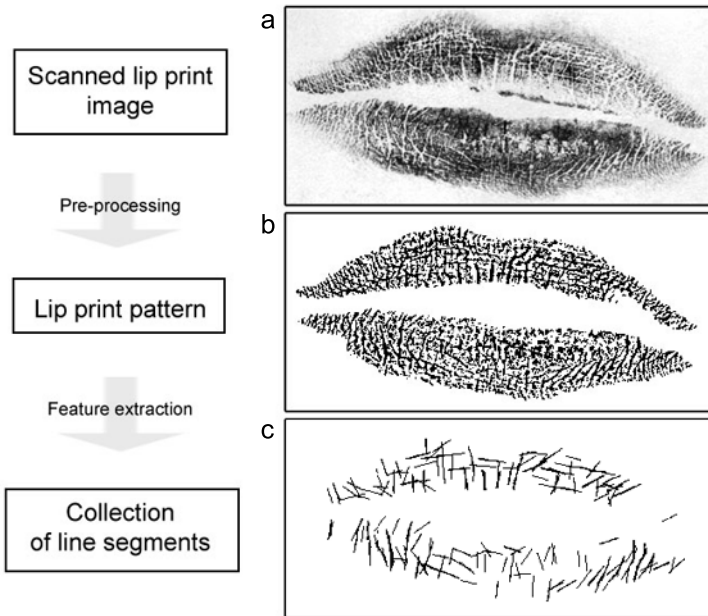


Fig. 2 Stages of lip print image processing.

The final stage of lip print analysis is segments detection. The aim of this procedure is approximation of lip pattern by a corresponding collection of segments, which form a set of characteristics of a lip print. Segments detection is based on lip pattern (Figure 2b) and a set of straight lines resulted from Hough transformation.

In the first step, straight lines resulted from Hough transform are transferred to Cartesian coordinate system:

$$\theta \in \left[0, \frac{\pi}{4}\right) \cup \left(\frac{3\pi}{4}, \pi\right) \Rightarrow \begin{cases} y = \text{round}(r - xctg\theta) \\ x = 1, 2, \dots, w \end{cases} \quad (7)$$

$$\theta \in \left[\frac{\pi}{4}, \frac{3\pi}{4}\right] \Rightarrow \begin{cases} x = \text{round}(r - yctg\theta) \\ y = 1, 2, \dots, h \end{cases} \quad (8)$$

Next, points which form a straight line and overlap the lip pattern are examined. For every straight line so-called segment detection array is created. If a point (x, y) of the straight line lies in the lip pattern area, then value of the corresponding coordinate in the segment detection array is set to 1.

In the last step, parameters of segments corresponding to lip pattern are determined. Analysis of the segment detection array for each of the straight lines is conducted. Beginning and end of each sequence of ones determines beginning and end coordinates of the corresponding segment, while number of consecutive ones defines its length. Graphical interpretation of a set of segments resulted from the segment detection procedure is presented of Figure 2c.

3 Comparison

The comparison of lip prints is conducted using the mean differences similarity measure. A chart describing the comparison procedure was presented in Figure 3. Mean differences were introduced by Czekanowski in [11, 13, 17] as an anthropological index describing variation within species. This method was used in studies of biometric personal identification based on handwritten signatures. The results obtained were characterized by a low classification error [10, 17].

As a result of feature extraction algorithms a collection of segments approximating lip pattern is obtained. A single segment can be described by the following features:

- $s^{\vartheta} = \sqrt{(x^e - x^b)^2 + (y^e - y^b)^2}$ - length of the segment,
- $s^{\alpha} = \arccos(|x^e - x^b|/s^{\vartheta})$ - angle between the segment and the X -axis,
- $s^{\xi} = (x^b + x^e)/2$ - x -coordinate of the segment's midpoint,
- $s^{\delta} = (y^b + y^e)/2$ - y -coordinate of the segment's midpoint.

Values x^b, y^b and x^e, y^e represent the coordinates of the beginning and end of the segment respectively.

Let there be two lip print images described by collections of segments: $S_1 = \{s_{1,1}, s_{1,2}, \dots, s_{1,n}\}$, $S_2 = \{s_{2,1}, s_{2,2}, \dots, s_{2,m}\}$. Let also $F = \{s^{\vartheta}, s^{\alpha}, s^{\xi}, s^{\delta}\}$ be a set containing characteristic features of a segment.

Let's denote $F_{1,i}^k$ the k -th feature of i -th segment of the set S_1 (where $1 \leq i \leq n$) and $F_{2,j}^k$ the k -th feature of j -th segment of the set S_2 (where where $1 \leq j \leq m$). To determine the effects of the k -th characteristic feature on the comparison results, a feature weight w_k satisfying the following condition was introduced:

- $w_k \in [0, 1]$,
- $\sum_{k=1}^{\text{card}(F)} w_k = 1$.

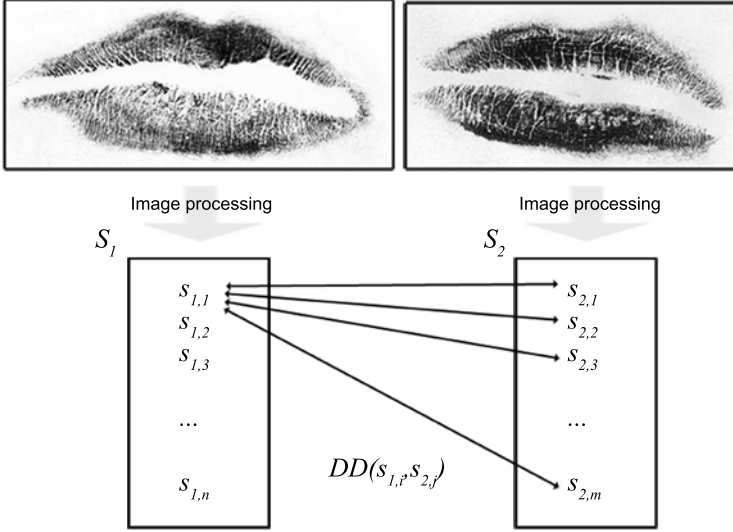


Fig. 3 Lip prints comparison procedure.

The weighted mean difference between two elements $s_{1,i} \in S_1$ and $s_{2,j} \in S_2$ (where $1 \leq i \leq n$, $1 \leq j \leq m$) is given by the formula:

$$DD(s_{1,i}, s_{2,j}) = 1 - \sum_{k=1}^{card(F)} |F_{1,i}^k - F_{2,j}^k| w_k \quad (9)$$

The mean differences similarity coefficient $sim_{DD}(S_1, S_2)$ is defined as follows:

$$sim_{DD}(S_1, S_2) = \frac{\sum_{i=1}^n \max(DD(s_{1,i}, s_{2,1}), \dots, DD(s_{1,i}, s_{2,m}))}{n} \quad (10)$$

Mean differences similarity coefficient defined in such a way takes values in the range of [0,1]. A value of 1 means that the compared images are identical. Mean differences similarity measure is asymmetric. This means that $sim_{DD}(S_1, S_2) \neq sim_{DD}(S_2, S_1)$. This property stems from the fact that the value n in the formula (10) is the number of segments the set S_1 , which is usually different than the number of elements of the set S_2 .

4 The Course and Results of the Studies

The studies were carried out for lip print images obtained from 30 people. The test database consisted of 120 images (4 lip prints for every individual). Performance

of the proposed lip print recognition method was tested for different distribution of weights applied to the segment characteristic features (step 0.05) and for similarity threshold values in the range of 0 to 1 (step 0.001).

All the lip print images were compared "round robin". Similarity of two images was calculated using the mean differences similarity measure presented in the previous chapter. Due to the asymmetry of the similarity measure every comparison was conducted back and forth. Similarity threshold value was adopted as a classification criterion. Figure 4a presents five weight distributions of segment characteristic features for which EER error took the lowest value.

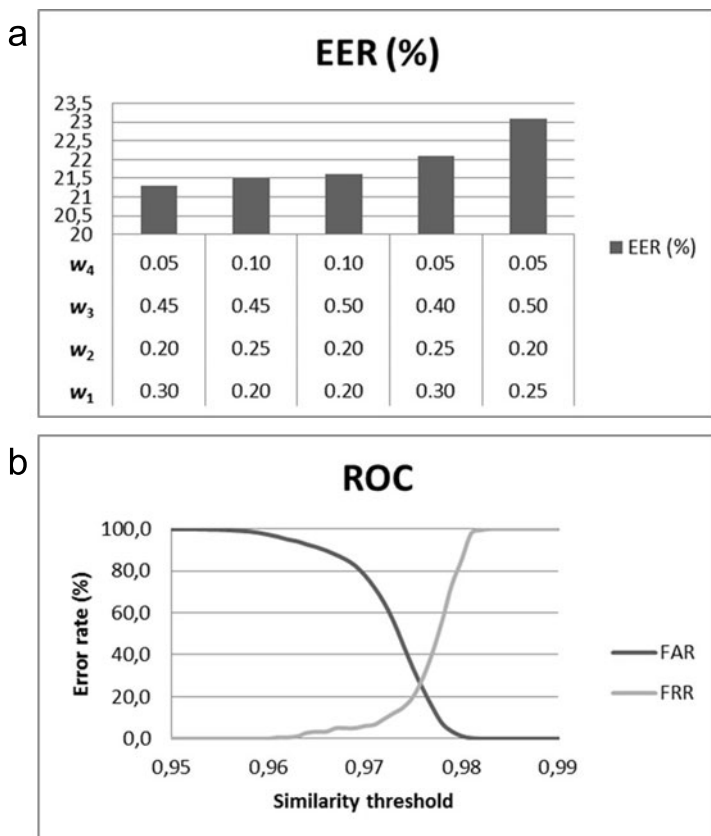


Fig. 4 Experimental results: a) EER values for different weights applied for the segment's characteristic features, b) ROC curves of FAR and FRR.

The lowest ERR error (21,3%) was obtained for the following weights: $w_1=0.30$, $w_2=0.20$, $w_3=0.45$, $w_4=0.05$. Analysis of test results shows that parameter s^f has the highest impact on the obtained results. Figure 4b presents the ROC curves for the optimal distribution of segment characteristic features.

It can be noticed that the proposed similarity measure differentiates the test lip prints in a small degree. To obtain satisfactory results, similarity parameter has to be calculated with a high accuracy (at least 3 decimal places).

5 Conclusions

Lip prints coming from the same person vary in shape. It is caused by the fact that human lips, in opposite to fingertips, are very ductile. Therefore, automatic identification of lip prints is a very difficult task. Even though there is a strong demand for such systems, this area of image recognition is still insufficiently investigated. There are no solutions performing lip prints recognition available publicly.

Presented experimental results show that the resulting error is average. However, the proposed algorithms are still improved and have a high potential. This makes the results obtained encouraging enough to allow further research in this direction. As part of further work, it is planned to enhance the pre-processing and feature extraction algorithms and test other similarity measures.

References

- [1] Campbell, B.M.: Similarity coefficients for classifying relevés. *Vegetatio* 37, 101–109 (1978)
- [2] Choras, M.: The lip as a biometric. *Pattern Analysis and Applications* 13(1), 105–112 (2010)
- [3] Czekanowski, J.: Zur differential-diagnose der Neandertalgruppe. *Korrespbl. Dt. Ges. Anthropol* 40, 44–47 (1909)
- [4] Duda, R., Hart, P.: Use of the Hough transformation to detect lines and curves in pictures. *Comm. ACM* 15 (1972)
- [5] Hough, P.V.C.: Method and means for recognizing complex pattern. U.S. Patent No. 3069654 (1962)
- [6] Illingworth, J., Kittler, J.: A survey of the Hough transform. *Computer Vision Graphics and Image Processing* 44(1), 87–116 (1988)
- [7] Johnston, J.W.: Similarity Indices I: What Do They Measure?, p. 68 (1976)
- [8] Kasprzak J., Leczyńska B.: Cheiloscopy - Human identification on the basis of lip trace, pp. 11–29 (2001) (in Polish)
- [9] Newton, M.: *The Encyclopedia of Crime Scene Investigation*, p. 42 (2008)
- [10] Porwik, P., Doroz, R., Wrobel, K.: A new signature similarity measure based on windows allocation technique. *Int. Journal of Computer Information Systems and Industrial Management Applications (IJCISIM)* 2, 297–305 (2010)
- [11] Sharma, P., Saxena, S., Rathod, V.: Comparative reliability of cheiloscopy and palatoscopy in human identification. *Indian Journal of Dental Research* 20(4), 453–457 (2009)
- [12] Siegel, J., Saukko, P., Knupfer, G.: *Encyclopedia of Forensic Science*, pp. 358–362 (2000)
- [13] Smacki, L.: Lip traces recognition based on lines pattern. *Journal of Medical Informatics and Technologies* 15, 53–60 (2010)

- [14] Smacki, L., Porwik, P., Tomaszycycki, K., Kwarcinska, S.: The lip print recognition using Hough transform. *Journal of Medical Informatics and Technologies* 14, 31–38 (2010)
- [15] Suzuki, K., Tsuchihashi, Y.: Personal identification by means of lip prints. *Journal of Forensic Medicine* 17, 52–57 (1970)
- [16] Tsuchihashi, Y.: Studies on personal identification by means of lip prints. *Forensic Science*, 127–231 (1974)
- [17] Wrobel, K., Doroz, R.: New signature similarity measure based on average differences. *Journal of Medical Informatics and Technologies* 12, 51–56 (2008)

Human Vein Pattern Correlation – A Comparison of Segmentation Methods

Rafał Kabaciński and Mateusz Kowalski

Abstract. In this paper two methods of human vein pattern segmentation from low quality images, called frequency high-pass filtration and local minima analysis, proposed by authors in their previous article are compared with the often used local thresholding algorithm. These methods are evaluated using results of classification performed by a correlation algorithm. Evaluation was carried out on 400 collected images, and shows that proposed methods are worth to consider in human vein pattern segmentation.

1 Introduction

Biometrics is a promising field in access control systems technology. It is very easy to use, does not require from users to remember passwords, and eliminates a need for additional devices like tokens or access cards. Among many human physiological features, the blood vessels pattern acquired from an image taken in the near infrared light (NIR) has many advantages. It is different for every human, it does not change with age, it can be acquired contactless, and requires blood in veins, to be registered, which solves the liveness test problem. Recently, due to a price drop of CCD sensors, and increasing computing power, this emerging technology is becoming more popular. The very important step in using vein pattern for recognition purposes is its extraction from background of the hand. Many researchers use local thresholding (LT) method at this stage. In previous paper [4] authors proposed two alternative methods and compared them with LT, using manually assigned bifurcation points of the blood vessel pattern. In this paper the results of a correlation based recognition algorithm are used for evaluation and comparison of these methods.

This paper is organized as follows. In Section 2 the database used in the evaluation process is described. Section 3 introduces the methods discussed in this paper.

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

The correlation classification algorithm is explained in Section 4. In Section 5 the comparison methodology is presented. Section 6 shows the results of evaluation, and Section 7 contains some concluding remarks and proposals for further research.

2 Dataset

To compare the effectiveness of segmentation methods described further in this paper, a dataset of appropriate images needed to be acquired. Near infrared images acquisition system designed by authors was used to collect such data. The system is described in [4].

The dataset consists of 400 images taken from 50 volunteers aged 22-25 years. For every volunteer four images of both the left and the right palms were registered one by one. In general we may say that the dataset consisted of images of 100 objects, 4 images for each one. During the acquisition process the free positioning system was used, which means that there were no elements holding or blocking a palm. That might cause translations and rotations of the vein pattern in the images taken from the same person.

Images have the resolution 768×1024 , however in process stages it was decreased to 240×320 to reduce the calculation cost. Three new datasets of binary images were obtained after processing the images from described dataset using three segmentation methods. The image function for each image was equal '1' for pixels corresponding to the vein pattern (in this paper called sometimes 'white') and '0' for background pixels ('black').

3 Segmentation Methods

3.1 Local Thresholding

Local thresholding is the most common segmentation method used by vein pattern researchers ([2], [5], [6]). In this method a threshold is being calculated for each pixel, basing on its local neighbourhood. In the version of LT algorithm discussed in this paper, mean value of each pixel surroundings was used as the threshold. This algorithm requires good filtration in preprocessing stage. In this paper median and Gauss filters are used, which means that 3 parameters are to be set. Next 2 parameters are used in the algorithm. To improve the output, the morphological erosion is applied, whose mask size is the last parameter. In total 6 parameters are needed to be chosen. This algorithm has high calculation cost because it requires to analyze the neighbourhood of each pixel.

3.2 DFT Filtration Method

The first proposed method applies Discrete Fourier Transformation (DFT) to perform filtration in frequency domain. As a structured element in the image, the vein

pattern should have visible representation in frequency components. By suppressing right frequencies, filtration will make vessels darker than their surroundings. In this paper the Gaussian filter is used to high-pass filtration in frequency domain. Segmentation process consists of the following steps:

1. Calculate the raw image spectrum.
2. Multiply the spectrum with Gaussian mask (Fig. 1b).
3. Calculate the new image with inverse DFT (Fig. 1c), without rescaling the image to 0-255 grayscale.
4. Smooth result image with the block filter (Fig. 1d,e).
5. Binarize the image by the global thresholding with threshold equal to 0.

The result image after thinning with the algorithm described in [3] is presented in Figure 1f. Parameters required by this method are two Gaussian mask parameters and the size of the block filter.

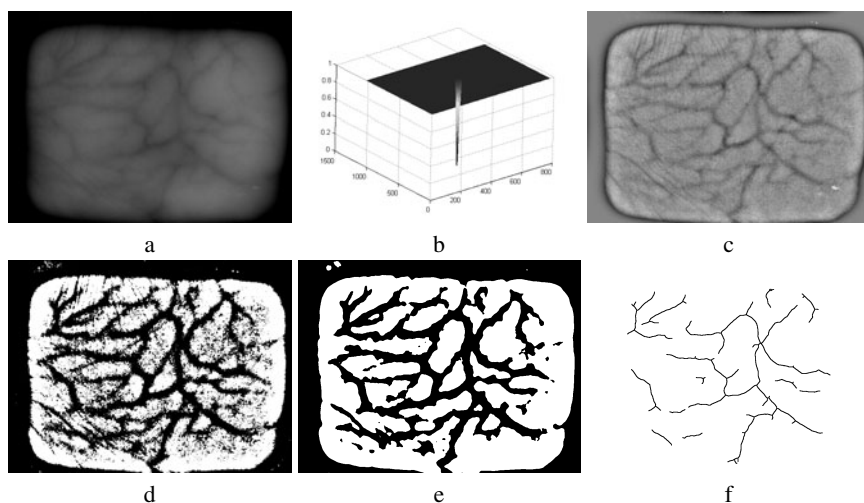


Fig. 1 Top row shows the frequency filtration process: a - input image, b - filter transfer function, c - filtration result in greyscale; bottom row shows postprocessing stages in inverted colors: d - frequency filtration result (without scaling to greyscale) global thresholded at zero, e - image d with average filtration for smoothing applied before thresholding, f - thinning results.

3.3 The Gradient-Based Segmentation Method

In NIR images blood vessels are darker than surrounding tissues, which means that they are local minima of the image function. This fact is utilised by LT method which finds pixels that are darker than the mean value of their neighbourhood. To reduce the calculation cost, local minima can be searched only in rows and columns, basing on the image function gradient. The drawback of this method is a need of a

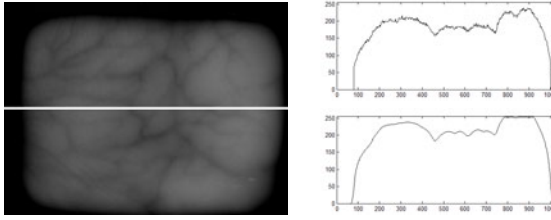


Fig. 2 On left original image, with marked row; on right the same row before filtration (on top) and after filtration, and scaling (bottom)

very good filtration which will entirely remove the noise without destroying vein pattern. The example image and its middle row profile before filtration is presented in Figure 2. In this paper a 2D low-pass Gaussian filter is used in the loop. Filtration is applied as long as every single pixel wide minimum in predefined rows and columns (in $1/4$, $1/2$, and $3/4$ of the image width and height) is removed, or the loop exceed a defined number of iterations. The example image and its middle row profile, after filtration is presented in Figure 2. Detection of local extrema is based on zero crossing points of a directional derivative of the image function. To calculate directional derivative only a value of a preceding pixel is needed. A result of the local minimum detection is presented in Figure 3. Local minima are represented by the set of separated points and many of them are a result of remaining noise, but points representing blood vessels are formed in curvilinear clusters. Because morphological closing of such image would cause many parasitic branches, after thinning and finding the local minimum 10 pixels are also set to 1. The result image can be seen in Figure 3. The morphological erosion is then used to eliminate lines isolated from clusters. Finally, morphological closing and second erosion are applied to improve the continuity and smoothness of the pattern skeleton, removing remaining smaller artefacts in the process. The image after thinning with algorithm described in [3], is presented in Figure 3. This method requires 5 parameters in total.

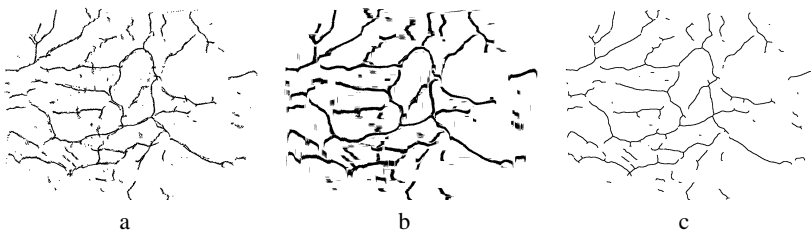


Fig. 3 Results of gradient method in inverted colors: a - local minima marked as dots, b - marked by lines and c - final result

4 Correlation

For every image obtained using one of the segmentation methods, the degree of similarity to all other images obtained using the same method was determined. Degree of similarity is defined basing on 2D correlation function, which is given by the formula:

$$R_{xy}(i, j) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} X(m, n)Y^*(m + i, n + j) \quad (1)$$

where: X and Y are images, m and n are image indices and i and j represents the translation of the second image.

For binary images like those discussed here 2D correlation function can be interpreted as follows:

1. On the second image pixels with value '1' i pixels in vertical direction and j pixels in horizontal direction are translated.
2. Number of pixels with value 1 placed in corresponding cells of matrices representing images is calculated (Pixels that are out of frame on second image in result of translation are omitted).

Degree of similarity (further in this article called DS) is defined as maximum value of 2D correlation function divided by the number of pixels with value 1 on a picture, on which the number of those pixels is smaller. Degree of similarity was calculated for images on which number of white pixels differed not more than two times. In this way, using correlation we solve the problem related to translations of the vein pattern on the image due to free positioning of a hand.

The crucial difference between images containing the vein pattern from the same person can also be result of rotations related to a free positioning of the hand. To reduce the influence of free positioning, the correlation was calculated repeatedly for the second image rotated around center of mass of the vein pattern from -15 degree to 15 degree with 1 degree step. In such solution it may happen that DS_{XY} (DS for images X and Y where Y is rotated) is not equal to DS_{YX} (DS for images Y and X where X is rotated). DS was calculated for binary images containing 8-pixels neighbourhood path corresponding to the vein pattern. For the same blood vessel from the same pattern route of path could have different deformation as a consequence of the thinning method. Therefore dilation with 9x9 pixels wide mask was applied to the path.

DSs obtained this way for different segmentation methods were gathered in Comparison Matrices and compared, what is described in the next Section.

5 Comparison of Methods

Results of comparison (values determining the DS) for each method were gathered in Comparison Matrix (called further CM). Images were put in order depending

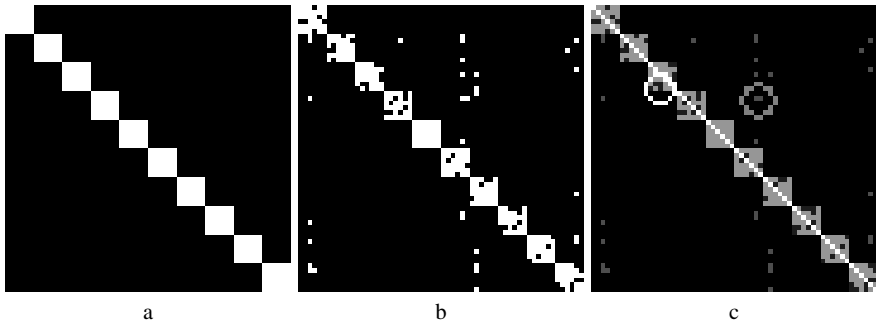


Fig. 4 Exemplary images of Comparison Matrices: a - optimal pattern of CM, b - CM for some threshold, c - visualization of TP(white, gray around diagonal), FN(black marked in white circle), TN(black), FP(gray on the black background marked in gray circle) sets as a result of comparison CM from (b) with optimal CM pattern from (a)

on a person and a hand from which they were taken. As an effect on Comparison Matrix diagonal results of comparison of each image with itself were placed and in a diagonal surrounding were results of comparison of images taken from the same object. Figure 4b presents optimal pattern of comparisons Matrix, white fields corresponds to DS value '1' for images taken from the same object, black fields corresponds to DS value '0' for images taken from different objects. Fields in real matrix of comparison contains obviously different values from range 0 to 1.

We say that two pictures X and Y are recognised as representing the same vein pattern if DS_{XY} calculated for them is greater than the chosen threshold. If we establish the threshold, we can divide Comparison Matrix into two sets of elements: equal to 1 for pair of images recognised as representing the same vein pattern and equal to 0 for pair of images recognised as representing different vein patterns. The goal is to find such a value for the threshold, if it is possible, that Comparison Matrix is most similar to the pattern presented in Figure 4b. Comparing CM for established threshold with mentioned pattern we can define:

1. TP - True Positive - set of pairs of images correctly recognised as representing the same vein pattern
2. FN - False Negative - set of pairs of images incorrectly recognised as representing different vein patterns (representing in fact the same vein pattern). Together TP and FN creates mentioned surrounding of CM diagonal presented in Fig. 4c.
3. TN - True Negative - pairs of images correctly recognised as representing different vein patterns.
4. FP - False Positive - pairs of images incorrectly recognised as representing the same vein pattern (representing in fact different vein patterns).

All sets TP, FN, TN and FP are presented in different colors in Figure 4. Comparison of mentioned segmentation methods was done basing on rates taken from ([1]):

- Genuine Acceptance Rate (GAR)

$$GAR = \frac{TP}{FN + TP} \times 100\% \tag{2}$$

- False Rejection Rate (FRR)

$$FRR = \frac{FN}{FN + TP} \times 100\% \tag{3}$$

- False Acceptance Rate (FAR)

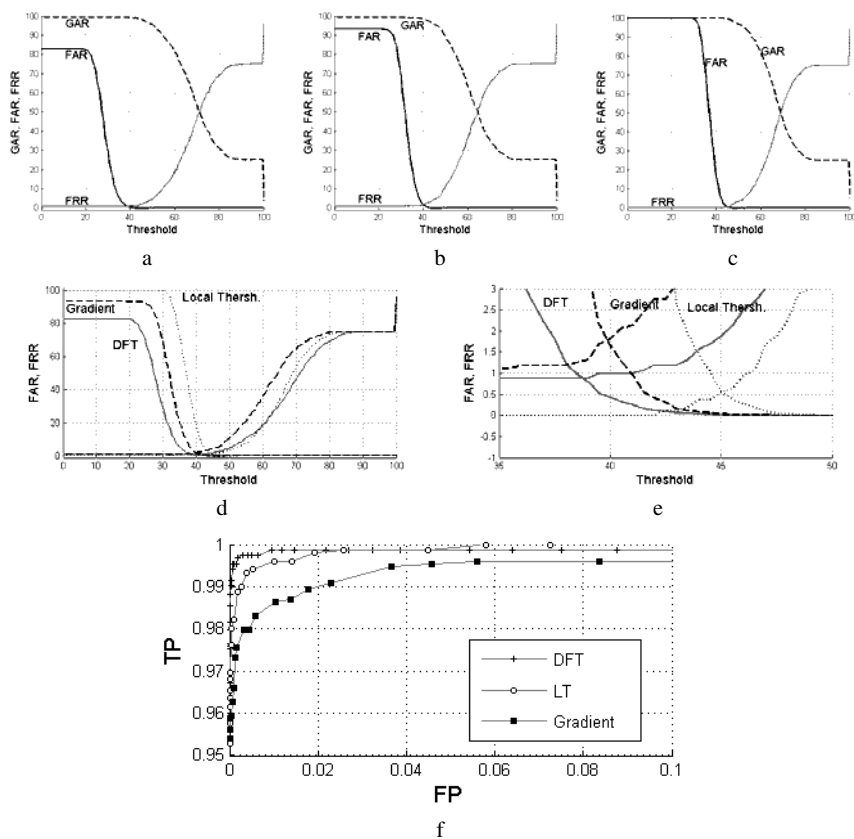


Fig. 5 Top row shows diagrams of FAR(black line), FRR(gray line), GAR(black dashed line) rates for each method of segmentation: a - DFT Filtration method, b - Gradient-based method and c - Local thresholding method; bottom row shows comparative diagram of FAR and FRR rates for all methods (d) and enlargement of this diagram around FAR and FRR crossing points(e), line styles correspond to following methods: DFT Filtration - gray line, Gradient-based - black dashed line, Local thresholding - black dotted line, finally ROC graph of compared methods is presented (f)

$$FAR = \frac{FP}{FP + TN} \times 100\% \quad (4)$$

In the next Section these rates are more deeply analysed.

6 Results

Results of preformed calculations are presented in Figure 5 as a diagrams of rates FAR (black line), GAR (black dashed line), FRR (gray line) for each method. To simplify analysis comparative diagram of FAR and FRR rates was made on one picture (line styles corresponds to: Fourier based method - gray line, Gradient based method - black dashed line, Local Thresholding - black dotted line). In the same Figure the image showing a diagram enlarged around crossing points of FAR and FRR plots is presented.

Results of these analysis are gathered in Table 1. Diagrams can be analysed in two ways. In the first analysis the threshold minimizing both FAR and FRR rates is searched. Then values of those rates for this threshold are compared. In this case results are comparable however Local thresholding method gives the best effect. The second analysis, more interesting from a practical point of view, consists in searching a threshold for which FAR rate is equal to the assumed value. Then FRR value is compared for this threshold. In these case two different values of FAR were considered ($FAR = 0,1$ and $FAR = 0,01$) and FRRs outcomes were compared for them. Again results were similar however the best effect was given by the DFT Filtration method. The same conclusion is true for Receiver operating characteristic (ROC) graph analysis.

Table 1 Hand vein pattern correlation results.

	Fourier	Gradient	Local thresh.
FAR and FRR crossing point			
Threshold	39	40	45
GAR	99,02	98,19	99,44
FRR	0,88	1,81	0,56
FAR	0,88	1,81	0,56
Threshold for $FAR \geq 0.1\%$			
Threshold	42	40	47
GAR	98,81	96,62	98,31
FRR	1,19	3,38	1,69
FAR	0,12	0,12	0,09
Threshold for $FAR \geq 0.01\%$			
Threshold	45,5	46,5	48,5
GAR	97,87	94,81	97,12
FRR	2,13	5,19	2,88
FAR	0,01	0,01	0,01

7 Conclusions

In this paper three methods of vein pattern segmentation from low-quality images are compared. One of them called Local Thresholding appears to be popular in texts about vein pattern classification. Two other methods are proposed by the authors and introduced in details in [4]. Images from dataset collected by the authors were processed with each of this methods and classified using the algorithm based on 2D correlation function. Results of classification were used to compare these methods.

Results of comparison shows that methods gives comparable effects. In this case some important advantages of proposed methods are worth noting. Local minima analysis based method has lower calculation and memory cost.

Frequency domain filtration method requires only three parameters to be established, therefore it is easier to use. Although this method has the biggest calculation cost, dedicated electronic circuits performing DFT are available. Therefore access control device based on this method can be faster than these based on LT method.

References

- [1] Anil, K.J., Flynn, P.J., Ross, A.A.: Handbook of biometrics. Springer, Berlin (2008)
- [2] Crisan, S., Tarnovan, I.G., Crisan, T.E.: Vein pattern recognition. Image enhancement and feature extraction algorithms. In: 15th Symposium on Novelty in Electrical Measurements and Instrumentation in parallel with the 12th Workshop on ADC Modelling and Testing, Iasi (2007)
- [3] Lam, L., Lee, S.-W., Suen Ch, Y.: IEEE Transactions on Pattern Analysis And Machine Intelligence 14, 869–885 (1992)
- [4] Kabaciński, R., Kowalski, M.: Advances in Soft Computing 84, 105–112 (2010)
- [5] Shahin, M., Badawi, A., Kamel, M.: International Journal of Biological and Life Sciences 2, 141–148 (2006)
- [6] Wang, L., Leedham, G., Cho, D.S.-Y.: Pattern Recognition 41, 920–929 (2008)

Knuckle Recognition for Human Identification

Michał Choraś and Rafał Kozik

Abstract. This paper is the continuation of our previous work - hereby we report new results proving the effectiveness of our knuckle recognition method based on texture features. We use Probabilistic Hough Transform and SURF features as well as the 3-step classification methodology. Hereby we present promising results achieved for recently published PolyU knuckle database.

1 Introduction

Despite the fact that several biometrics modalities (fingerprints, face, hand geometry and iris) have already been deployed in large scale systems many drawbacks of the mentioned modalities are still reported. Requirements from users and stakeholders prove that biometric systems will soon have to be user-centric meaning requirements-free, friendly, accepted, trusted and mobile [1][2]. Therefore new modalities are still extensively investigated.

One of emerging modalities satisfying such requirements is knuckle also termed as FKP (finger-knuckle-print) [3][4].

It has several advantages, such as:

- invariance to emotions and other behavioral aspects,
- high-textured regions,
- up to 5 distinctive biometrics samples per one hand,
- it is easily accessible.

Knuckle biometrics methods can be used in biometric systems for user-centric, contactless and unrestricted access control e.g. for medium-security access control or verification systems dedicated for mobile devices (e.g. smartphones and mobile telecommunication services) [5].

The samples of knuckles images (index and middle finger for left hand and middle finger for right hand respectively) from PolyU Database are presented in Figure 1[6].

Michał Choraś · Rafał Kozik

Image Processing Group, Institute of Telecommunications, UT & LS Bydgoszcz

e-mail: chorasm@utp.edu.pl

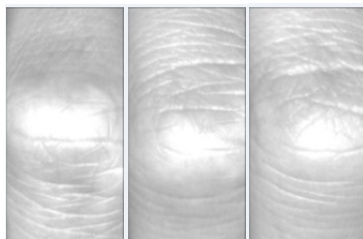


Fig. 1 Sample knuckle images from PolyU Database for one person (index and middle finger for left hand and middle finger for right hand [6]).

Even though knuckle biometrics is relatively unknown and new modality, there already are some results and feature extraction methods. So called KnuckleCodes have been proposed and other well known feature extraction methods such as DCT, PCA, LDA, ICA, orientation phase, Gabor filters have been investigated showing very good identification results [3][7][8][9][10][11].

Hereby, we propose to use texture feature extraction methods such as Probabilistic Hough Transform (PHT) and Speeded Up Robust Features (SURF). In this work we present our methodology and provide further verification of our system using PolyU knuckle database [6].

The overview of our knuckle recognition methodology is given in Section 2. Lines extraction algorithm is presented in Section 3. The proposed feature extraction methodology is presented in Section 4. Three-step classification methodology, experiments and results are described in Section 5.

2 Overview of Knuckle Recognition Methodology

In this paper we propose knuckle recognition methodology which allows to reduce query time significantly. The diagram explaining the proposed methodology is shown in Figure 2.

Firstly, the knuckle image is obtained from individual requesting access to the system. The knuckle image is preliminary processed to gain the characteristic features. The preprocessing consists of edge detection and thresholding.

The image is further analyzed by means of Probabilistic Hough Transform (PHT), which allows to extract basic feature vector. Then the database is queried in order to retrieve N matching samples. For that purpose we use KD-Tree in order to obtain results even faster. For each N samples in the resulting set, the complex feature vectors are compared. The approach with kNN allows to decrease the complex computation without losing the overall system effectiveness (as discussed in details in section 5). Finally in order to eliminate false positives, SURF descriptor is used to for M nearest matches obtained in the previous step.

The results prove (both of IIT and PolyU databases) that this methodology is working well for knuckles due to their intrinsic properties.

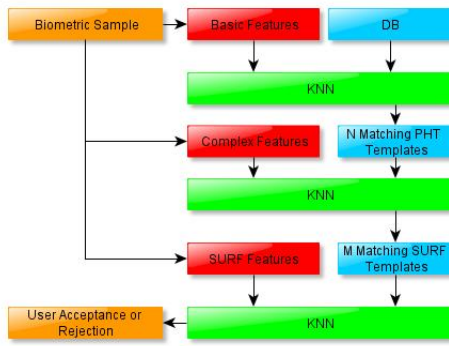


Fig. 2 Knuckle recognition methodology overview

3 Lines Extraction

In our feature extraction method, we focus mainly on lines and wrinkles located on bending area of finger joints. Therefore the image is first segmented (via thresholding) using an adaptive threshold estimated by means of equation (1):

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

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

The result of adaptive thresholding can be seen in Fig. 3 B. It can be noticed that such an image is quite noisy, since some edges suffer from line discontinuities while the background is filled with small spots.

This problem is solved by adapting the Probabilistic Hough Transform (PHT). Later, the PHT also is used in our approach for extracting the dominant orientation and building the "basic feature vector".

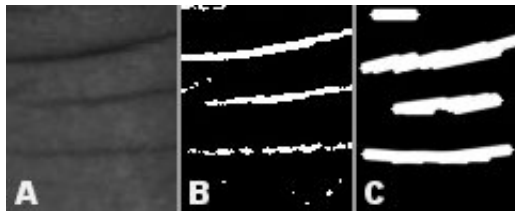


Fig. 3 The knuckle image example form IIT Database (A), enhanced major lines after thresholding (B), and the lines detected by PHT (C)

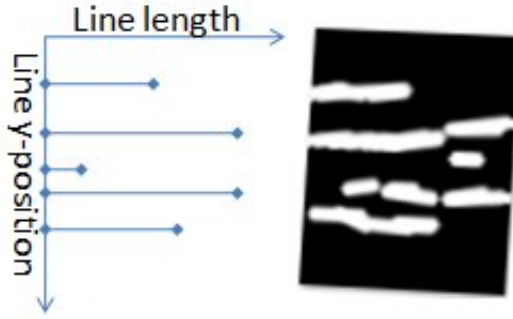


Fig. 4 The basic feature vector is built using the PHT output and the 1D histogram

4 Feature Extraction

4.1 Basic Features

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

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

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

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

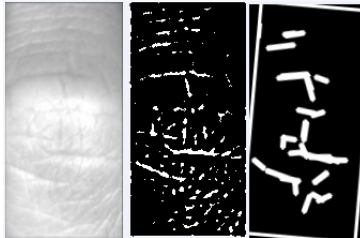


Fig. 5 Sample knuckle images and their representation after applying PHT transform

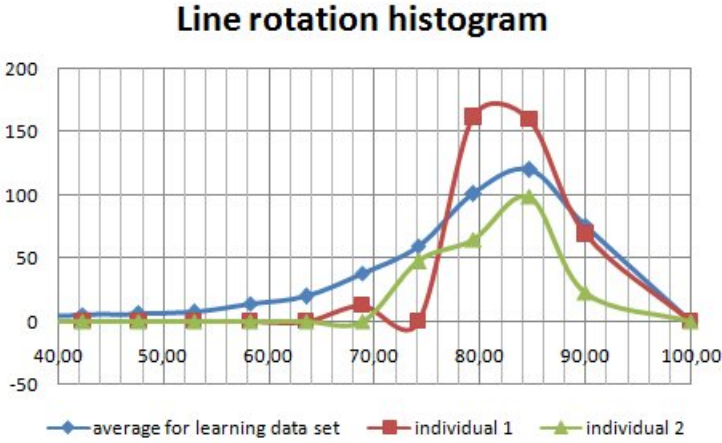


Fig. 6 Line rotation histogram. The x-axis indicates the line rotation while the y-axis indicates the cumulative lines length.

is used to build the feature vector. The 30-bins 1D histogram is adapted as it is shown in Fig. 4. Such an approach is based on the fact that the longest and characteristic lines of knuckle are concentrated around one rotation angle (as proved in Fig. 6).

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

4.2 Complex Features

The set of line descriptors (eq. 2) obtained from Hough transform are converted to image representation giving input for matching algorithm as it is shown in Fig. 7.

Both query and template images (chosen from kNN selected by basic feature classifier) are transformed and compared using the Euclidean metrics.

The output of matching block is the scoring map, which is 30x30 of size. The size is determined by searching ranges. In this case the template image is offset in $\langle -15, 15 \rangle$ range both on x and y dimension as is it defined by formula 3, where i and j define the element in scoring map, H and W define query image width and height, while q and t represent query and template images, respectively.

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

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

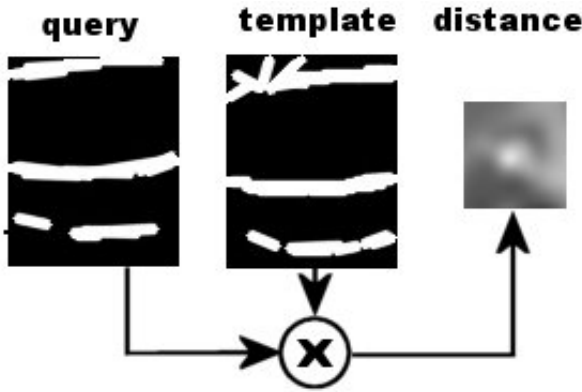


Fig. 7 During the procedure of PHT models matching a map of distances is generated and the closest match is chosen.

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

The 5 images from kNN set yielding the lowest score are chosen giving the input for SURF-based classifier described in section 4.3.

4.3 SURF Features

The SURF stands for Speeded Up Robust Features and is robust image detector and descriptor. It was firstly presented by Herbert Bay in 2006 [13]. It is widely used in object recognition and 3D reconstruction. The key-point of the SURF detector is the determinant of the Hessian matrix, which is the matrix (eq. 4) of partial derivative of the luminance function.

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

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

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

If the determinant value is greater than the threshold (estimated during experiments on learning data set) then it is considered as fiducial point. The greater the

threshold is, the less points (but "strong" ones) are detected. For each of the fiducial points the texture descriptor is calculated.

In our approach we use the SURF points to find the closest matching (if any) between querying image and the templates selected by PHT-based classifier. Firstly, the points yielding the Hessian determinant value greater than threshold are selected for both querying and template images resulting in two points data set. Basing on texture descriptors the matching pairs between those sets are found and the outliers (points in one data set that do not have representative in the second data set) are removed. Then the matching cost between those sets is estimated using eq. 6:

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

where N , d , p and q represents the number of matching pairs, Euclidean distance, point from template image and point from query image respectively. Example of such a mapping is shown in Fig. 8:

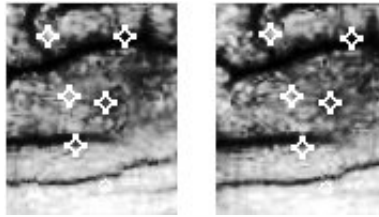


Fig. 8 Detected fiducial SURF points for queering image and its corresponding matches for the template image

5 Experiments and Results

For experiments the proposed classification methodology steps were configured as follows:

1. selecting 50 images on the basis of basic vector,
2. selecting 5 images on the basis of PHT feature vector,
3. selecting 1 closest match using SURF descriptor.

When basic feature vector for particular knuckle image is computed it is looked up in data base to find k nearest neighbors yielding the nearest Euclidean distance. The k number was determined empirically as an compromise of system effectiveness and system performance.

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

The proposed approach was tested using PolyU Knuckle Database [6]. The knuckle images were obtained from 165 individuals. Each individual contributed more 10-12 image samples per left index, left middle, right middle and right index finger.

For efficiency assessment the classic 10-fold method was applied and average of experiments results is presented. The average equal error rate obtained during experiments is 2.42%. Test were repeated for left middle, right middle and right index fingers. The ROC curve presented in Figure 9 shows that each finger gives comparable results.

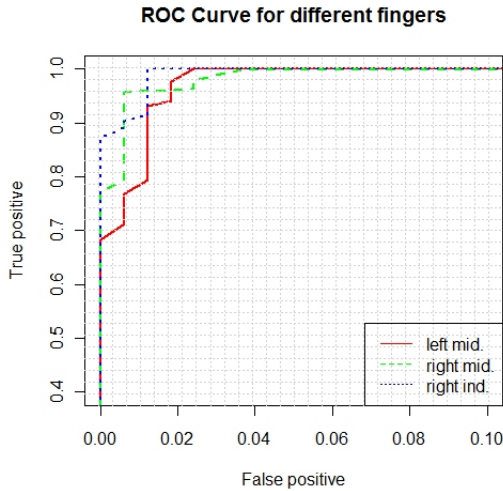


Fig. 9 ROC curves for different fingers - middle finger (both left and right) and right index finger

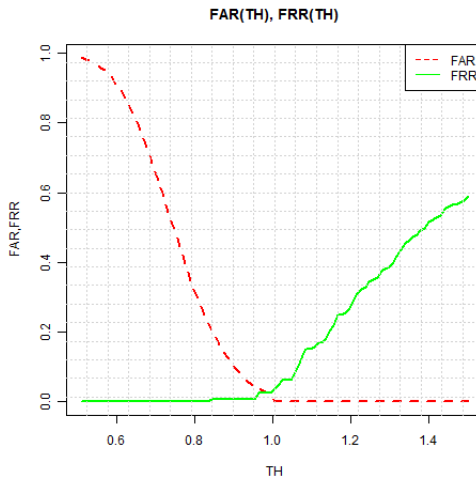


Fig. 10 FAR vs. FRR for right middle finger

The FAR and FRR vs. system threshold for one of the experiments is shown in Figure 10.

The experiments show that combination of PHT and SURF gives better results than each of this method used separately.

The SURF failed so often due to the fact it was unable to find matching between query knuckle image and the template. Those recognition errors were covered by PHT. However, PHT failed when it came to distinct two or more similar knuckles in k nearest neighbors. In this situation SURF was more accurate.

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

6 Conclusions

In this work further developments in human identification based on knuckle texture features are presented. The major contribution of the paper are: new knuckle feature extraction methodology based on PHT (Probabilistic Hough Transform) and SURF features as well as original classification methodology. Our system and the achieved results have been also discussed. Our work proves that knuckle should be considered as promising emerging modality.

We reported comparable promising results for PolyU knuckle database as in our previous work on IIT Delhi database [14], [15]. In contrary to our previous work, we additionally measured effectiveness of our method for different fingers.

References

- [1] Fratric, I., Ribaric, S.: Real-Time Model-Based Hand Localization for Unsupervised Palmar Image Acquisition. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 1280–1289. Springer, Heidelberg (2009)
- [2] Methani, C., Namboodiri, A.M.: Pose Invariant Palmprint Recognition. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 577–586. Springer, Heidelberg (2009)
- [3] Morales, A., Ferrer, M.A., Travieso, C.M., Alonso, J.B.: A knuckles texture verification method in a transformed domain. In: Proc. of 1st Spanish Workshop on Biometrics (on CD), Girona, Spain (2007)
- [4] Zhang, L., Zhang, L., Zhang, D.: Finger-knuckle-print: a new biometric identifier. In: Proceedings of the IEEE International Conference on Image Processing (2009)
- [5] Kozik, R., Choras, M.: Combined Shape and Texture Information for Palmprint Biometrics. *Journal of Information Assurance and Security* 5(1), 58–63 (2010)
- [6] <http://www4.comp.polyu.edu.hk/biometrics/FKP.htm>
- [7] Kumar, A., Zhou, Y.: Human Identification using Knuckle Codes. In: Proc. BTAS (2009)
- [8] Kumar, A., Ravikanth, C.: Personal authentication using finger knuckle surface. *IEEE Trans. Information Forensics and Security* 4(1), 98–110 (2009)
- [9] Kumar, A., Zhou, Y.: Personal identification using finger knuckle orientation features. *Electronics Letters* 45(20) (September 2009)

- [10] Zhang, L., Zhang, L., Zhang, D., Hailong Zhu, H.: Online Finger-Knuckle-Print Verification for Personal Authentication. *Pattern Recognition* 43(7), 2560–2571 (2010)
- [11] Zhang, L., Zhang, L., Zhang, D.: Finger-knuckle-print verification based on band-limited phase-only correlation. In: *Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns*, pp. 141–148 (2009)
- [12] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001*, vol. 1, pp. 511–518 (2001)
- [13] Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
- [14] Choraś, K.R.: Knuckle Biometrics Based on Texture Features. In: *Proc. of International Workshop on Emerging Techniques and Challenges for Hand-based Biometrics (ETCHB 2010)*, Stambul, August 2010. IEEE CS Press, Los Alamitos (2010)
- [15] Choraś, K.R.: Knuckle Biometrics for Human Identification. In: Choras (ed.) *Image Processing and Communication Challenges 2*. ASC, pp. 95–102. Springer, Heidelberg (2010)

Part II
FEATURES, LEARNING AND
CLASSIFIERS

Modified Ranked Order Adaptive Median Filter for Impulse Noise Removal

Anna Fabijańska

Abstract. In this paper problem of impulse noise removal is considered. Specifically, modifications of ranked order adaptive median filter (RAMF) are proposed. RAMF is popular, well established and effective switching median filter for denoising images corrupted by impulse noise. However, the modifications proposed in this paper significantly improve its results, especially in case of highly corrupted images. Results of denosing of images under a wide range of noise corruption (5-95%) using the original and the modified ranked order median filter are presented, compared and discussed. Comparison is made by means of PSNR and SSIM index.

1 Introduction

Median filtration is the most popular method for impulse noise removal. Applied to grayscale images is a neighborhood intensity-ranking algorithm. Each pixel in the output image equals to the median intensity of its closest neighbors in the original image [1, 2]. The size of the neighborhood is determined by a filtering window, which passes through the image pixel-by-pixel [4, 10].

Regardless its popularity, there are drawbacks to median filtration. It is performed uniformly across the image and filters both: noisy and uncorrupted pixels. In consequence details are compromised and image looks blurry [6]. Therefore, many modifications of classical median filtration have been proposed. They involve different techniques for noise determination in order to decrease the number of pixels subjected to median filtration. In consequence impulse noise is suppressed while preserving image details [3, 7, 8, 9, 11, 14, 15].

One of the well established methods for impulse noise removal is ranked order adaptive median filter (RAMF) [7]. It incorporates ranked-order statistics to detect impulse noise. Although RAMF is very robust in noise detection it falters during noise removal, specifically in case images under a high level of noise corruption. In

Anna Fabijańska

Computer Engineering Department, Technical University of Lodz

18/22 Stefanowskiego Str., 90-924 Lodz, Poland

e-mail: an_fab@kis.p.lodz.pl

this paper improvements for RAMF are proposed. They allow to improve efficiency of image restoration.

2 Noise Model

Impulse noise is considered [5]. It affects the pixel x_{ij} at location (i, j) in accordance to probability density function given by Equation [1]

$$f(x_{ij}) = \begin{cases} p_a & \text{for } x_{ij} = a \\ 1 - p & \text{for } x_{ij} = y_{ij} \\ p_b & \text{for } x_{ij} = b \end{cases} \quad (1)$$

where: a and b are minimum (black dot) and maximum (white dot) of intensity, p_a and p_b denote probability of intensity a and p_b respectively, $p = p_a + p_b$ is noise density and y_{ij} is an intensity of the pixel at location (i, j) in the original image. If either p_a or p_b is zero the impulse noise is called unipolar noise. If neither probability is zero and especially if they are equal, impulse noise is called bipolar noise or salt-and-pepper noise [5].

3 Ranked Order Adaptive Median Filter

Ranked order adaptive median filter (RAMF) is well established filter for impulse noise removal. It was proposed by Hwang and Haddad in [7]. Here only brief review of the filter is given. For further details please follow the source paper. RAMF is switching median filter i.e. only pixels identified as noisy are subjected to median filtration. The remaining pixels are left unchanged. In order to detect noisy pixels ranked-order-differences are used. During RAMF performance each pixel and its neighborhood are tested twice. First test, checks if there are residual impulses in the median filter output. In this step two statistics T_- and T_+ are applied. They are given by Equation [2]

$$\begin{cases} T_- = x_{med} - x_{min} \\ T_+ = x_{med} - x_{max} \end{cases} \quad (2)$$

where x_{med} , x_{max} and x_{min} denote median, maximum and minimum of intensity within the filtering window W_D respectively.

Possible values of statistics T_- and T_+ are presented in Table [1]. In cases (a), (b) and (c) size of filtering window is increased and first level test is repeated until case (d) is true or maximum window size is reached. The maximum size of the filtering window W_D depends on noise density p . Hwang and Haddad claim that $W_D = 5$ is sufficient for $p = 0.3$ while $W_D = 11$ is needed for noise density $p = 0.7$. In case (d) the second test in the pixel neighborhood is performed. This time statistics U_- and U_+ (Eq. [3]) are used.

$$\begin{cases} U_- = x_{ij} - x_{min} \\ U_+ = x_{ij} - x_{max} \end{cases} \quad (3)$$

Possible values of statistics U_- and U_+ are presented in Table II. If (a), (b) or (c) then the pixel is considered noisy and RAMF output is the x_{med} from the previous step. Otherwise pixel is left unchanged.

Table 1 Possible values of pair of statistics (T_- , T_+) and (U_- , U_+)

case	T_-/U_-	T_+/U_+
(a)	0	0
(b)	0	negative
(c)	positive	0
(d)	positive	negative

4 Modified Ranked Order Median Filter

In its source version RAMF is very robust in impulse noise detection but significantly less efficient in noise correction. It is mainly because a new intensity of noisy pixel is determined in relation to its neighborhood without regarding if neighboring pixels are noisy or not. Moreover, before RAMF is performed noise density p must be estimated in order to determine maximum mask size used for noise detection. The modified ranked order adaptive median filter (MRAMF) is a remedy for the abovementioned weaknesses of the source method. The proposed MRAMF works in three stages which are:

- impulse noise detection;
- noise density estimation;
- noise correction.

The main changes in comparison to the original method are: introduction of binary decision map (BDM), simplified noise density estimation scheme and recursive filtering with adaptive mask size. They are briefly characterised below.

Impulse noise detection

Robust noise detection scheme utilized by RAMF is also used by its modified version proposed in this paper. However, while RAMF immediately subjects pixels identified noisy to median filtration, MRAMF firstly remembers position of all noisy pixels recognized in the image by marking their location on binary decision map. Specifically, 1's on the binary decision map define locations of pixels identified corrupted and 0's are assigned to the uncorrupted pixels. BDM is next used in the following steps for image denosing. It indicates pixels which should be subjected to median filtration and pixels which should be excluded when retrieving original pixel intensity. Moreover, in opposite to the source method, constant-size window is used by MRAMF during noise detection. Specifically, $W_D = 11$ regardless noise density p .

Noise density estimation

Impulse noise detection scheme utilized by RAMF is very robust therefore noise density p estimation was simplified in modified ranked order median filter. Value of p is determined based on number of detected noisy pixels. Hence, 1's on binary decision map are simply counted and their number is divided by a total number of pixels in the considered image in accordance with Equation 4

$$p = \left[\sum_{i=1}^M \sum_{j=1}^N \frac{BDM(x_{ij})}{MN} \right] 100\% \quad (4)$$

where $BDM(x_{ij})$ denotes value on binary decision map corresponding with location (i, j) and M, N are image dimensions.

Noise correction

In the last step noise removal is carried out. Switching median filter is applied for image denoising. Filtering is performed based on binary decision map. The map indicates both: pixels whose intensity should be reconstructed and pixels which should be excluded when retrieving original pixel intensity.

Specifically, only pixels identified as corrupted by noise (indicated by 1's on binary decision map) are subjected to median filtration. Their new value is median of uncorrupted pixels (indicated by 0's on binary decision map) located in the window W_D . The current pixel is excluded from the median filtration as it is identified as corrupted by noise. This concept is presented in Figure 1

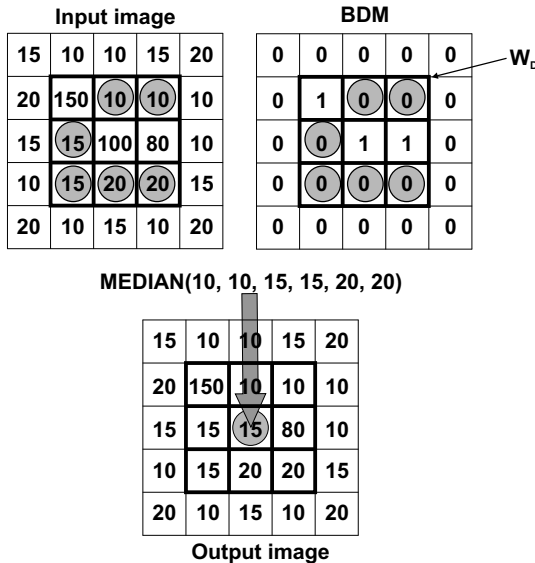


Fig. 1 Idea of median filtration using binary decision map.

Dimensions of filtering window change in function of the noise density in the way given in Table 2. Dimensions of filtering window were determined experimentally.

Table 2 Window size in function of noise density

noise density	$W_D \times W_D$
$0\% < p \leq 10\%$	3×3
$10\% < p \leq 30\%$	5×5
$30\% < p \leq 50\%$	7×7
$50\% < p \leq 70\%$	9×9
$p > 70\%$	11×11

5 Results

5.1 Image Fidelity Measures

The comparison of noise removal quality by the considered algorithms is made by means of peak signal-to-noise ratio (PSNR) [12] and structural similarity (SSIM) index [12, 13] given by Equation 5 and 7 respectively.

$$PSNR = 10 \log_{10} \left(\frac{L^2}{MSE} \right) dB \quad (5)$$

where L denotes dynamic range of intensities and MSE is mean squared error given by Equation 6.

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (y_{ij} - x_{ij})^2 \quad (6)$$

where M, N are image dimensions, y_{ij} denote intensity of pixel at location (i, j) in the original image and x_{ij} denote intensity of pixel at location (i, j) in the filtered image.

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (7)$$

where x, y are pixel coordinates, $l(x, y)$ is luminance component given by Equation 8, $c(x, y)$ is contrast component given by Equation 9, $s(x, y)$ is structure component given by Equation 10 and α, β, γ are positive parameters used to adjust the importance of the components.

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (8)$$

where μ_y, μ_x are mean intensities of the original and the restored image respectively and C_1 is constant equal to $(K_1L)^2, K_1 \ll 1$.

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (9)$$

where σ_y, σ_x are standard deviations of intensities in the original and the restored image respectively and C_2 is constant equal to $(K_2L)^2, K_2 \ll 1$.

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (10)$$

where σ_{xy} is cross-correlation between the original and restored image and C_3 is constant.

In accordance to suggestions given in [13] the following values of parameters were used: $\alpha = \beta = \gamma = 1, K_1 = 0.01, K_2 = 0.03$ and $C_3 = C_2/2$. Local SSIM indexes were computed using window size of 11×11 pixels passing throughout the image pixel-by-pixel. Local indexes were next averaged in order to compute global SSIM index. In case of ideal match of two images SSIM index equals to 1. Then, it decreases to 0 together with decreasing visual similarity between compared images.

5.2 Results and Discussion

The proposed MRAMF was extensively tested on different images corrupted by salt and pepper noise. Images under a wide range of noise corruption (from 5% to 95%) were considered. In this section results of applying RAMF and MRAMF to standard test images are presented. Specifically, images of *Bridge* and *Elaine* (see Fig. 2) are used. They are 8-bit monochromatic images of the spatial resolution 512×512 pixels.



Fig. 2 Standard test images: (a) Bridge, (b) Elaine.

Results of impulse noise removal from images of *Elaine* and *Bridge* under a wide range of noise corruption are presented in Figures 3 and 4 respectively. First column presents noisy image. The noise density p is indicated on the left. Second column

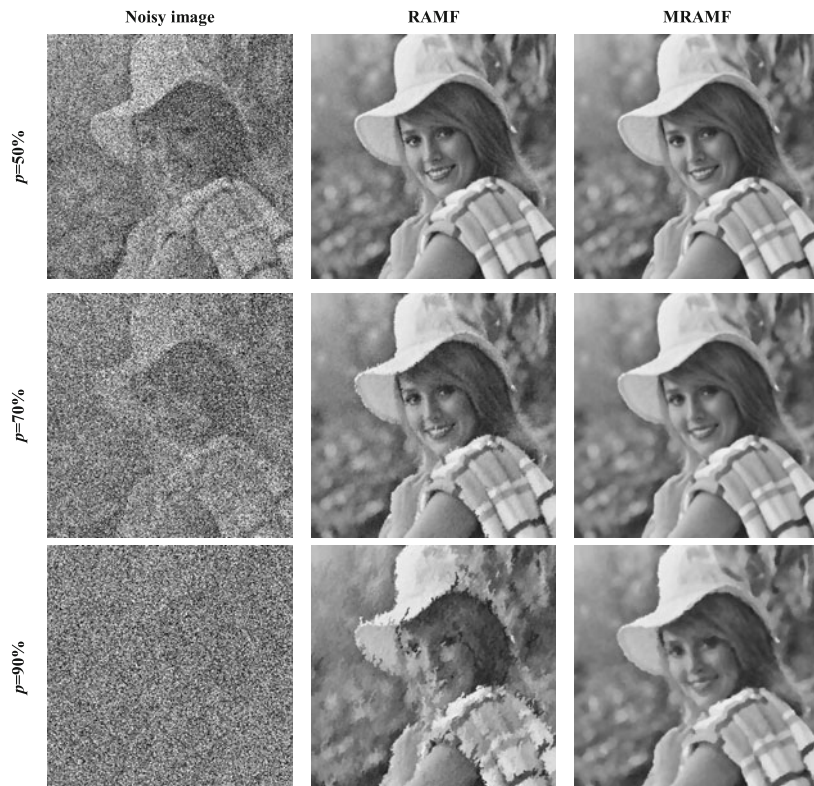


Fig. 3 Results of noise removal from image of Elaine.

presents results of image denoising using RAMF filter. In the third column images restored by using the introduced MRAMF filter can be seen.

The comparison of noise removal quality by means of PSNR and SSIM index is presented in Figures 5a and 5b respectively. Again, images of *Elaine* and *Bridge* are considered. The name of the image is indicated on each graph.

Results presented in Figures 3, 4 and 5 clearly show, that the proposed modified ranked order adaptive median filter (MRAMF) significantly outperforms the source method i.e. RAMF [7]. Specifically, in case of high level of noise corruption superiority of the proposed method can be seen. It is proven by higher values of both: PSNR and SSIM index. Moreover, for a wide range of noise corruption SSIM indexes of images recovered by MRAMF are always higher than those of ranked-order median filter in its source version. This means that the proposed approach is more robust in preserving details and ensures more visually pleasing output.

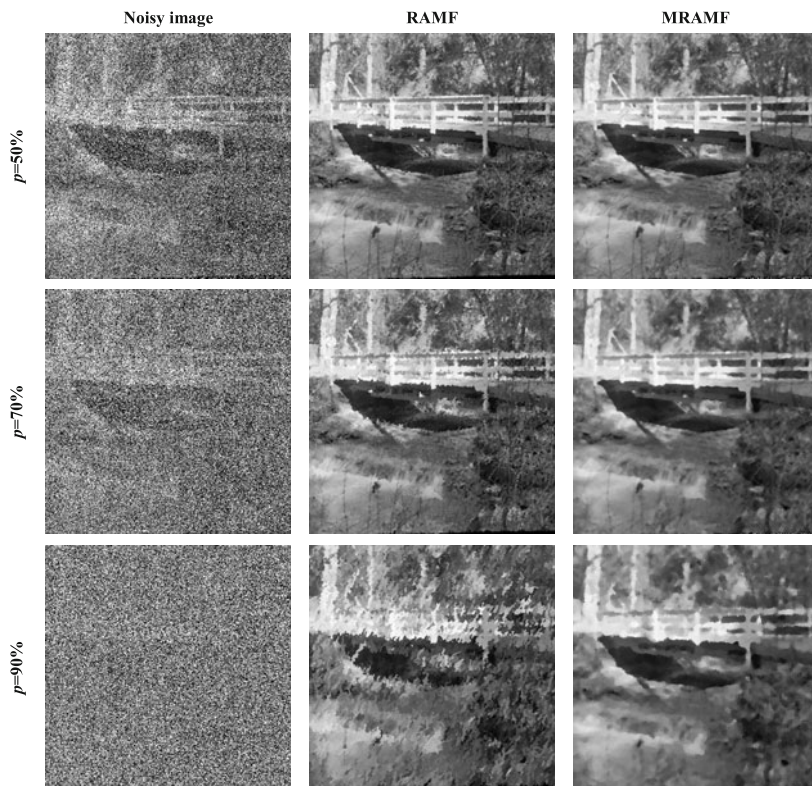


Fig. 4 Results of noise removal from image of Bridge.

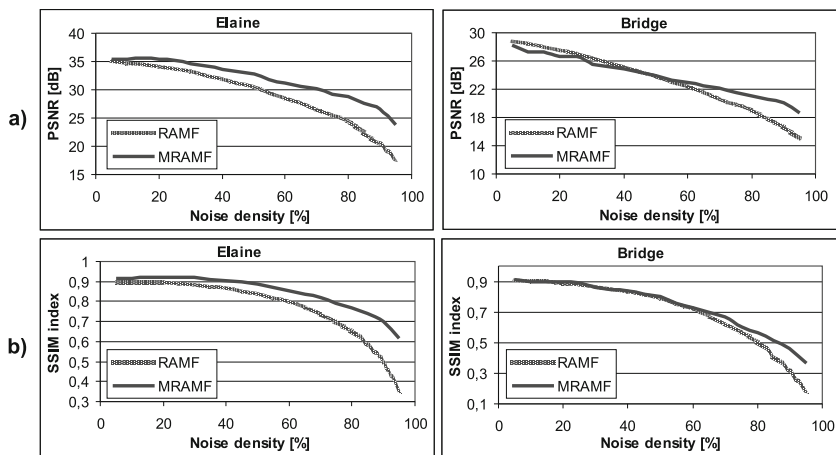


Fig. 5 Comparison of noise removal quality: (a) PSNR, (b) SSIM index.

6 Conclusions

In this paper problem of impulse noise removal was considered. Specifically, modification of well established ranked order adaptive median filter (RAMF) was proposed. The introduced method (i.e. MRAMF) fixes main weaknesses of the source method and increases quality of image restoration.

Results of extensive simulations were presented and compared with those obtained by the source method. Analysis of obtained results reveals that for a wide range of noise corruption (5%-95%) MRAMF provides high quality results comprehensible in overall image contents. Moreover, the new algorithm can be successfully applied for restoration of extremely-corrupted images.

Acknowledgement. The presented work is supported by the Ministry of Science and Higher Education of Poland from funds for science in years 2010-2012 as a research project no. N N516 490439.

References

- [1] Arce, G.R., Gallagher, N.C., Nides, T.: Median filters: theory and applications. In: Huang, T. (ed.) *Advances in Computer Vision and Image Processing*. JAI Press, CT (1986)
- [2] Astola, J., Kuosmanen, P.: *Fundamentals of nonlinear digital filtering*. CRC Press, USA (1997)
- [3] Chen, T., Wu, H.R.: Adaptive impulse detection using center-weighted median filters. *IEEE Signal Processing Letters* 8(1), 1–3 (2001)
- [4] Gil, J., Werman, M.: Computing 2-D min, median, and max filters. *IEEE Trans. Pattern Analysis and Machine Intelligence* 15(5), 504–507 (1993)
- [5] Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Prentice Hall, USA (2007)
- [6] Guangjin, Z., Jiegu, L.: Some problems of 2D morphological and median filters. *Journal of Shanghai University (English Edition)* 1(3), 245–248 (1997)
- [7] Hwang, H., Haddad, R.A.: Adaptive median filters: New algorithms and results. *IEEE Trans. Image Processing* 4(4), 499–502 (1995)
- [8] Lin, H.M., Wilson, A.N.: Median filters with adaptive length. *IEEE Trans. Circuits and Systems* 35(6), 675–690 (1988)
- [9] Ng, P.-E., Ma, K.-K.: A switching median filter with boundary discriminative noise detection for extremely corrupted images. *IEEE Trans. Image Processing* 15(6), 1506–1516 (2006)
- [10] Pitas, I., Venetsanopoulos, A.N.: Order statistics in digital image processing. *Proc. IEEE* 80(12), 1893–1921 (1992)
- [11] Vijaykumar, V.R., Vanathi, P.T., Kanagasabapathy, P., Ebenezer, D.: High density impulse noise removal using robust estimation based filter. *IAENG International Journal of Computer Science* 35(3) (2008), http://www.iaeng.org/IJCS/issues_v35/issue_3/index.html (accessed November 2010)
- [12] Wang, Z., Bovik, A.C.: Mean squared error: love it or leave it. *IEEE Signal Proc. Magazine* 26(1), 98–117 (2009)

- [13] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing* 13(4), 600–612 (2004)
- [14] Yin, L., Yang, R., Gabbouj, M., Neuvo, Y.: Weighted median filters: a tutorial. *IEEE Trans. Circuits and Systems* 43(3), 157–192 (1996)
- [15] Zvonarev, P.S., Apalkov, I.V., Khryashchev, V.V., Reznikova, I.V.: Neural network adaptive switching median filter for the restoration of impulse noise corrupted images. In: Kamel, M.S., Campilho, A.C. (eds.) *ICIAR 2005. LNCS*, vol. 3656, pp. 223–230. Springer, Heidelberg (2005)

Conceptual Design with the Use of Graph-Based Structures

Grażyna Ślusarczyk

Abstract. The paper deals with using a visual language and graph-based structures to support the conceptual phase of designing. The design ontology-based visual language is composed of design diagrams. A symbolic representation of design structures in terms of ontological objects and relations between them is defined, and then mapped into diagrams by a given realization. Diagrams are automatically transformed into the corresponding graph-based data structures. The knowledge stored in the graph representation of diagrams is transformed into first-order logic formulas and allows the design supporting system to reason about created designs. The approach is illustrated by examples of designing transmission towers.

1 Introduction

This paper deals with using knowledge to support the conceptual phase of designing by a computer system. Contemporary specialized CAD tools make it possible to replace sketches done at the beginning of the design process by computer drawings, which contain only general ideas about configurations of projects elements [8]. Such design drawings can be treated as elements of a problem-oriented visual language characteristic for a given design domain and used for communication between the user and the design system.

Supporting the conceptual phase of visual design by computer systems requires the appropriate problem-oriented design knowledge representation. In the presented approach design drawings, called diagrams, have their internal representations in the form of attributed hierarchical hypergraphs [4, 7]. Hyperedges of these hypergraphs represent diagram components and multi-argument relations among specified fragments of the components. Hypergraphs nested in hierarchical hyperedges represent arrangements of subparts of the diagram components.

Grażyna Ślusarczyk

The Faculty of Physics, Astronomy and Applied Computer Science

Jagiellonian University, Reymonta 4, 30-059 Kraków, Poland

e-mail: gslusarc@uj.edu.pl

The knowledge stored in the internal representation of diagrams enables the design supporting system to reason about designs. This knowledge is translated into ontology-based first-order logic sentences describing designs. Then, these sentences are compared with formulas expressing design criteria and potential incompatibilities with design objectives are reported.

The presented approach is illustrated by examples of designing transmission towers. Optimum design of a transmission tower can be considered at topological, geometrical and component level [1]. The optimization of transmission towers has been widely researched but mainly on the level of parametrical optimization with topology and geometry given a priori [6]. The problem of their topological optimization was considered in [5]. Transmission towers designed using the diagram language discussed here, with layouts represented in the form of hierarchical hypergraphs and created according to the given constraints (e.g. width of the tower), can constitute an initial population of evolutionary methods used in topological optimization.

2 Visual Languages

A proposed visual language is composed of design diagrams, which are finite configurations of geometric primitives and describe design solutions. A design diagram can be seen as a simplified engineering drawing and is treated as a visual abstraction representing a whole class of design solutions. Having specified type of design tasks, structures of designs, which provide a basis for creating design diagrams, are defined. At first, a design ontology being a classification and categorization of the knowledge [2] concerning the considered design domain is specified.

The design ontology is defined as $K = (O, O_R, A, Att, E, R, T)$, where:

- O is a set of objects, which can constitute elements of design problem solutions,
- O_R is a set of active objects, which are specified fragments of other objects and constitute arguments of relations,
- A is a set of attributes representing object properties, with a function Att which assigns sets of attributes to design objects of O ,
- E is a set of types with a given taxonomy, which classify objects of O ,
- R is a set of multi-argument relations, which can take place among active objects of O_R ,
- T is a set of operations which design objects can undergo.

By assigning domain-specific items to objects and relations of the design ontology the domain-oriented ontology, used for example to design transmission towers, is obtained.

Example 1. The design ontology used for designing transmission towers is specified as follows. The set O contains objects such as bottom, middle and top segments, trusses, truss panels, insulators and truss nodes. Truss nodes represent welding points and constitute active objects of other objects. The set $E = \{segment, truss, truss_panel, truss_node, insulator\}$ contains types which classify objects of

O. To each segment, truss, truss panel and insulator all its active objects of the type *truss_node* are assigned. To all objects the attributes *width*, *length* and *position* are assigned. To objects of the type *truss* and *truss_panel* the attributes *type* and *material* are also assigned.

The set of relations, $R = \{con, cnts\}$, contains a multi-argument relation *con* determining connections between truss nodes, and a multi-argument relation *cnts*, which holds among tower segments or trusses and their component panels. Operations performed on objects of a transmission tower enable the designer to divide segments into truss panels, trusses, and insulators connected to them, and divide trusses into their component panels.

The design ontology allows for defining structures of all designs of the considered domain of application. A design structure $Str = (Q, Att_Q, R_Q)$ is a set of relations among elements of the chosen set of attributed objects. Design structures are transformed with the use of a generative system composed of syntactic design rules defined over K , which relate to each operation of T changes of relations between designs objects that result from performing this operation and determine attribution of newly created objects. A design process is usually started from an initial design structure. Then operations of T are applied to objects of the structure transforming it into a new one.

Example 2. The structure of a transmission tower corresponding to the diagram presented in Fig 1b contains five objects of the type *truss_panel*, each with four objects of the type *truss_node*, three objects of the type *truss* representing horizontal trusses and three objects of the type *insulator*, each with two objects of the type *truss_node* (denoted in figures as black dots). The structure contains eleven connection relations between truss nodes.

A diagram shown in Fig 1c corresponds to an initial design structure of a transmission tower. It is composed of three objects corresponding to the bottom (Sg_b), middle (Sg_m) and top (Sg_t) segments, and eight active objects representing nodes in which truss panels of segments will be welded. It contains also four connecting relations between segment nodes. In successive design steps the segments are divided into trusses, truss panels and insulators. A design diagram shown in Fig 1b is obtained as a result of dividing the bottom segment into two truss panels (Te_1, Te_2), replacing the middle segment by one truss panel (Te_3), and dividing the top segment into two truss panels (Te_4, Te_5), three trusses ($T_1 - T_3$) and three insulators ($I_1 - I_3$).

A set *Snt* of all structures that can be derived from the initial structure by applying to objects operations of T and changing relations among objects according to the syntactic design rules associated with these operations, is called the syntax of a design language.

In order to create a diagram representing a design solution, the meaning must be given to objects of a design structure and relations among them. Thus, a family of realization functions τ_{Snt} , which maps design structures of the language syntax *Snt* into a domain of design diagrams, is defined. The domain of design diagrams contains a vocabulary S , which is composed of basic shapes corresponding to

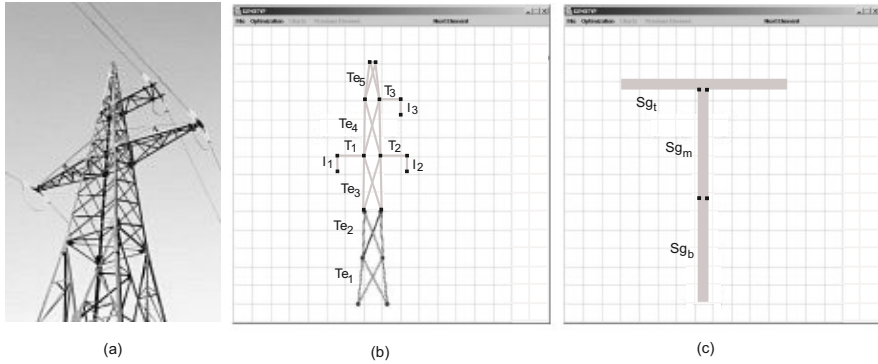


Fig. 1 a) A transmission tower b) a design diagram representing this tower c) an initial topology of the tower

objects specified in the design ontology, a set F of admissible transformations for elements of a vocabulary, and a set X of multi-argument geometric relations which hold among diagram components being transformed elements of a vocabulary.

Example 3. In case of designing transmission towers a vocabulary S of the design diagram domain contains shapes corresponding to different types of truss panels, line segments corresponding to horizontal trusses and insulators, and dots representing truss nodes. A set F of admissible transformations is composed of translations, rotations and scaling. A relation *touch*, which holds in a diagram if its components touch each other, denotes the connections among truss nodes.

Each realization function specifies an intended interpretation of a given design structure in a domain of design diagrams. For a given design structure Str composed of relations R_Q which hold among elements of the chosen set of attributed objects Q , a realization function is a pair $\tau = (\tau_1, \tau_2)$, where τ_1 assigns a pair $(s, f) \in S \times F$ composed of a visual primitive and the transformation which it should undergo to each object of Q , and τ_2 assigns one geometric relation of X to each relation of R_Q .

A generic design diagram Dg corresponding to the design structure Str and compatible with the realization specified by τ , is a sum of shapes of S under their corresponding transformations defined by τ_1 . Geometric relations that occur among diagram components and correspond to the relations of the structure Str are determined by τ_2 . A visual design language is composed of generic design diagrams being visualizations of design structures under the given realizations.

Example 4. Design diagrams representing transmission towers consist of two-dimensional elements corresponding to truss panels, horizontal segments representing horizontal trusses and of vertical or oblique segments corresponding to insulators, under which cables are suspended. The design diagram shown in Fig. 1b represents the front view of the transmission tower depicted in Fig. 1a.

3 Graph-Based Structures

Design diagrams are represented as attributed hierarchical layout hypergraphs [3, 7]. Hypergraphs encoding knowledge about diagrams constitute graph-based data structures, which can be automatically processed by a computer system and used for reasoning about design features. Hypergraphs enable us to describe relations between objects on different levels of hierarchy.

Layout hypergraphs contain two types of labelled hyperedges. Hyperedges of the first type, called object hyperedges, correspond to design objects of O represented by diagram components. Hyperedges of the second type, called relational hyperedges, represent multi-argument relations of R among active objects of O_R being fragments of design objects represented by object hyperedges. Hypergraph nodes correspond to active design objects of O_R , which can be arguments of relations. Hyperedges of the layout hypergraph are labelled by elements of O and R , while hypergraph nodes are labelled by elements of O_R . To each hyperedge of a hypergraph a sequence of different target nodes is assigned. Target nodes of an object hyperedge represent active objects assigned to the object which is represented by this hyperedge. Object hyperedges can contain nested hypergraphs which represent arrangements of parts of the object represented by a hierarchical hyperedge.

To represent characteristic features of objects attributing of object hyperedges and nodes is used. Values of hypergraph attributes allow for testing whether design requirements are met by the corresponding design diagrams.

Example 5. A hypergraph which represents the diagram from Fig. 1b is presented in Fig. 2. This hypergraph contains three hierarchical hyperedges labelled Sg_b , Sg_m ,

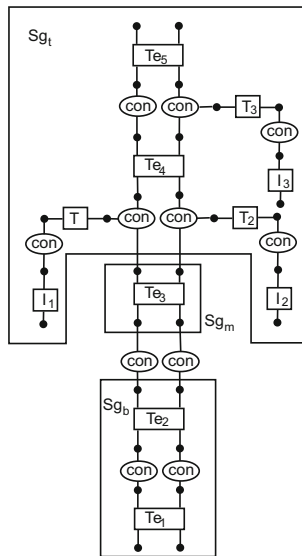


Fig. 2 A hierarchical hypergraph representing the design diagram from Fig. 1b

Sg_t and representing three segments of the tower. Hyperedge Sg_b contains two object hyperedges representing truss panels, while hyperedge Sg_m contains only one such hyperedge. Hyperedge Sg_t contains two object hyperedges representing truss panels, three object hyperedges representing horizontal trusses and three object hyperedges representing insulators. Hypergraph nodes represent truss nodes. Eleven relational hyperedges represent connections among trusses and between trusses and insulators. Attributes *width*, *height*, *position*, *type* and *material* are assigned to object hyperedges.

Design actions on diagrams are coupled with changes of their representations caused by operations applied on hierarchical layout hypergraphs [4, 7]. The *hyperedge development* operation defined in [3] corresponds to the division of an object into smaller parts. The *hyperedge suppression* operation defined in [4] allows one to redesign the chosen component of a diagram by first removing the existing division of a given object.

4 Reasoning from Internal Representations and Optimization of Designs

The system which supports the conceptual design phase assists the designer by reasoning about diagrams on the basis of their graph-based internal representations. It tests the compatibility of the solution with the specified constraints.

For a given ontology K , a vocabulary Σ , which consists of a set C of constant symbols representing objects of O_R and a set P of relation symbols representing relations of R , is defined. Constant symbols and variables constitute terms. For a set A of attributes, a set of functions which assign attributes to terms, is defined.

An atomic formula has the form $p(t_1, \dots, t_n)$, $t_1 = t_2$ or $w(t_1, \dots, t_n)$ where t_1, \dots, t_n are terms, $p \in P$ is an n -ary relation symbol and w is a constraint specified for attributes of terms t_1, \dots, t_n . Let F denote a set of general logical formulas built over atomic formulas using logical connectives, negation and quantifiers, which is consistent and closed under the consequence relation. Elements of F give a characterization of general properties of design diagrams. For example the sentence $\forall t \in truss \ width(t) < 5$ means that none of the horizontal trusses of a tower is longer than 5 meters.

The semantics is given to the formulas of F by a Σ -relational structure which consists of a domain of individuals and a way of associating with each of the elements of the vocabulary corresponding entities over the domain [2]. As design diagrams are represented by hypergraphs, the domain of the used relational structure is the set of hypergraph nodes and object hyperedges. Hypergraph nodes representing active objects are assigned to constant symbols and relational hyperedges are assigned to relation symbols. Then, a valuation on a Σ -relational structure, which assigns elements of the structure domain to variables, is defined.

Let Ψ be a set of formulas which represent design requirements. Let F be a set of formulas defined over a set of atomic formulas B satisfied by a given diagram Dg . A formula $\varphi \in \Psi$ is satisfied by a given diagram if there exists an atomic formula

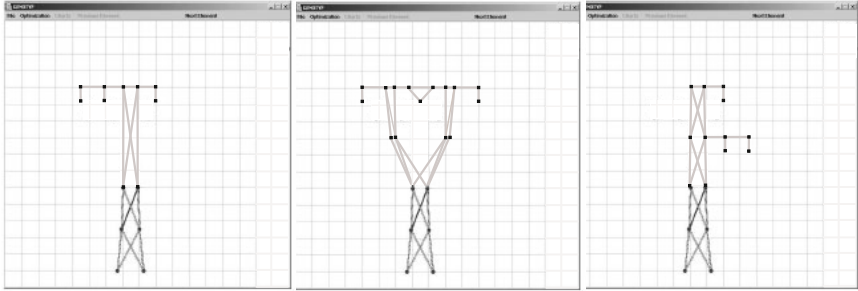


Fig. 3 Three layouts of transmission towers belonging to the initial population

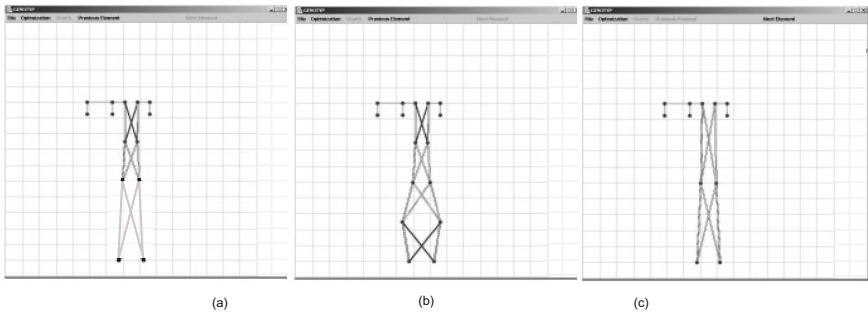


Fig. 4 a, b) Transmission towers of the first population c) a tower obtained after 20 steps of evolution

of B equivalent to φ or such a formula $\phi \in \mathbf{F}$ with an appropriate valuation can be found that φ can be inferred from ϕ .

Example 6. In case of designing transmission towers the design support system tests if the generated design is not too high. For a diagram presented in Fig. 1b the formula $height(t_1) + \dots + height(t_5) < 27$ is satisfied by a valuation $t_i = Te_i, i = 1, \dots, 5$, as $height(Te_1) = height(Te_2) = 5, height(Te_3) = height(Te_4) = 6, height(Te_5) = 4$. Analogously the system tests the width of the generated tower. The system can also compute distribution of stresses in the designed tower on the basis of cross-sectional areas of bars.

When layouts of transmission towers are coded in the form of hierarchical hypergraphs evolutionary methods with genetic operators modified to act on graphs can be used to optimize the towers' topology [1, 5]. The first step of optimization finds an optimal layout of a transmission tower, i.e., the one with the best positions of suspending the cables. The initial population is restricted here to towers with three places for power cables. Three diagrams of transmission towers, which belong to the initial population and are designed using our specialized editor, are shown in Fig. 3. After a predefined number of evolution steps a best fitted topology is selected.

The first step of optimization can be omitted when the designer knows from the engineering practice how the best structure to be used should look like, for example that it should have the *Y*-shaped layout. In the second step topology of the selected layout of the tower is optimized. Assuming that the first layout from Fig. 3 was chosen to be optimized, two elements of the first population are shown in Fig. 4a and Fig. 4b, while the best fitted structure received after 20 steps of evolution is shown in Fig. 4c.

5 Conclusions

This paper can be seen as the next step towards developing graph-based data structures being internal representations of design diagrams. Attributed hypergraph representations of designs allow the system to store knowledge about designs and serve as a base for reasoning about designs. Moreover, it seems that evolutionary programming together with a graph-based data representation is a powerful tool in the design process. Future applications will concern visual languages for designing three-dimensional forms of buildings. The hypergraph grammars which allow for automatic generation of hypergraph structures will be defined.

References

- [1] Borkowski, A., Grabska, E., Nikodem, P., Strug, B.: Searching for Innovative Structural Layouts by Means of Graph Grammars and Evolutionary Optimization. In: Proc. 2nd Int. Structural Eng. and Constr. Conf., Rome, pp. 475–480 (2003)
- [2] Coyne, R.D., Rosenman, M.A., Radeford, A.D., Balachandran, M., Gero, J.S.: Knowledge based design systems. Addison-Wesley Publishing Company, Reading (1990)
- [3] Grabska, E., Lachwa, A., Ślusarczyk, G., Grzesiak-Kopec, K., Lembas, J.: Hierarchical Layout Hypergraph Operations and Diagrammatic Reasoning. *Machine Graphics and Vision* 16, 23–38 (2007)
- [4] Grabska, E., Ślusarczyk, G., Le, T.L.: Visual Design and Reasoning with the Use of Hypergraph Transformations. In: Proc. of the Seventh International Workshop GT-VMT 2008, Budapest, vol. 10, pp. 305–318. EASST (2008)
- [5] Nikodem, P., Strug, B.: Graph Transformations in Evolutionary Design. In: Rutkowski, L., Siekmann, J.H., Tadeusiewicz, R., Zadeh, L.A. (eds.) ICAISC 2004. LNCS (LNAI), vol. 3070, pp. 456–461. Springer, Heidelberg (2004)
- [6] Save, M., Prager, W.: *Structural Optimization*. Plenum Press, New York (1990)
- [7] Ślusarczyk, G.: Hierarchical Hypergraph Transformations in Engineering Design. *Journal of Applied Computer Science* 11, 67–82 (2003)
- [8] Ware, C.: *Visual Thinking for Design*. Elsevier, Amsterdam (2008)

Estimation of Tolerance Relation on the Basis of Pairwise Comparisons

Leszek Klukowski

Abstract. The methods of estimation of the tolerance relation (overlapping partition) in a finite set on the basis of multiple pairwise comparisons with random errors – developed by the author - are discussed in the paper. Two types of comparisons are considered. The first type (binary) answers the question whether a pair of elements belongs to intersection of two or more subsets. The second type (multivalent) expresses the number of subsets of intersection comprising a pair. The estimates of the relation are determined on the basis of an appropriate discrete programming task. Two estimators are considered: the first one minimizes the sum of differences between relation form and comparisons. The second estimator rests on differences between relation form and medians from comparisons of each pair. The properties of the estimators are based on probabilistic inequalities and simulations.

1 Introduction

The tolerance relation (overlapping partition) divides a finite set of elements into family of subsets with at least one non-empty intersection (see e.g. [3], Ch. 5). The elements belonging to the same subset can be treated as items having a common feature. The relation is a model of many real-life phenomena, e.g. consanguinity in a group of descendants including brothers, sisters, stepbrothers and stepsisters; intersections comprise descendants of common parent. The relation can be estimated on the basis of pairwise comparisons with random errors. Such comparisons may be obtained on the basis of statistical tests or other decision procedures with similar features. Two forms of comparisons are considered in the paper: binary and multivalent. Any binary comparison evaluates existence of common feature of elements from a pair. Any multivalent comparison evaluates number of common features in a pair.

Leszek Klukowski
Systems Research Institute Polish Academy of Sciences,
6 Newelska Str., 01-447 Warsaw, Poland
e-mail: Leszek.Klukowski@ibspan.waw.pl

The purpose of the paper is presentation the estimators of the relation, developed by the author (see [7], [9], [6]) and their properties. The estimators rest on the idea of nearest adjoining order ([10], [2]) and require weak assumptions about distributions of comparison errors. Main assumptions comprise: unimodal distributions with mode and median equal zero and independent comparisons of each pair in the case of $N > 1$; the form of distributions may be unknown (nonparametric approach). The assumptions may be verified with the use of statistical tests. An example of application of the binary estimator - to partitioning a set of functions expressing profitability of treasury securities - is presented in [6]. The similar approach has been applied to estimation of the preference relation (2007b).

The problem of estimation of the tolerance relation have been discussed in the literature, see e.g. [3], but without formal properties of the estimates.

An example of estimation aimed at analysis of shapes of functions expressing profitability of treasury securities has been presented in [6].

The paper consists of six sections. The second section presents definitions and notations, the third problem formulation. Next sections, the fourth and fifth, describe two types of estimators, i.e. based on binary comparisons and multivalent comparisons. Last section concludes the results.

2 Basic Definitions and Notations

The tolerance relation divides the set $\mathbf{X} = \{x_1, \dots, x_m\}$ ($m \geq 3$) into n ($n \geq 2$) subsets χ_r^* ($r = 1, \dots, n$) with at least one non-empty intersection, i.e. the relation is not transitive. It satisfies the conditions:

$$\mathbf{X} = \bigcup_{r=1}^n \chi_r^* \quad \text{and there exists at least one pair of subsets} \\ \chi_r^* \quad \text{and} \quad \chi_s^* \quad (s \neq r) \quad \text{such that: } \chi_r^* \cap \chi_s^* \neq \mathbf{0}. \quad (1)$$

The tolerance relation is characterized with the use of the functions $T_1 : \mathbf{X} \times \mathbf{X} \rightarrow D_1$, $D_1 = \{0, 1\}$ or $T_2 : \mathbf{X} \times \mathbf{X} \rightarrow D_2$, $D_2 = \{0, 1, \dots, n\}$, defined as follows:

$$T_1(x_i, x_j) = \begin{cases} 0 & \text{if there exists } \chi_q^* \text{ such, that } (x_i, x_j) \in \chi_q^*, i \neq j; \\ 1 & \text{otherwise.} \end{cases} \quad (2)$$

$$T_2(x_i, x_j) = \#(\Omega_i^* \cap \Omega_j^*), \quad (3)$$

where:

$$\Omega_r^* = \{s \mid x_i \in \chi_s^*\},$$

$\#(\Xi)$ – the power (number of elements) of the set Ξ .

It is assumed, that the functions $T_f(\cdot)$ ($f = 1, 2$) determine the relation completely, i.e. there exists one-to-one relationship between the relation form and the functions $T_1(\cdot)$, $T_2(\cdot)$. The requirement is satisfied, for example, when each subset χ_q^* includes an element x_i , which is not included in any other subset χ_s^* ($s \neq q$) (i.e. $x_i \in \chi_q^*$ and $x_i \notin \chi_s^*$).

The problem corresponding to the first function $T_1(\cdot)$ is named binary approach, the second – $T_2(\cdot)$ – multivalent approach. The value $T_1(x_i, x_j)$ answers the question whether two elements (x_i, x_j) belong to the same subset χ_q^* ($1 \leq q \leq n$) (have any common feature), the value $T_2(x_i, x_j)$ – determines the number of subsets of an intersection including both elements (number of common features).

The multivalent approach can be extended by using additional function $T_3(x_i, x_j)$ defined as follows:

$$T_3(x_i, x_j) = \#(\Psi_i^* \cap \Psi_j^*), \quad (4)$$

where:

$$\Psi_r^* = \{1, \dots, n\} - \Omega_r^*.$$

The function $T_3(\cdot)$ expresses number of missing features in the pair (x_i, x_j) – taking into account the features existing in the set \mathbf{X} – and has to satisfy the same “completeness condition”, as the functions $T_1(\cdot)$, $T_2(\cdot)$.

The functions $T_1(\cdot)$, $T_2(\cdot)$, $T_3(\cdot)$ describe similarities and dissimilarities of elements from the set \mathbf{X} – on the basis of classical theory of sets, without fuzzy formalization.

3 Problem Formulation

The problem of estimation of the tolerance relation can be stated as follows. There exists the tolerance relation $\chi_1^*, \dots, \chi_n^*$ in a finite set \mathbf{X} . The relation is to be estimated on the basis of multiple pairwise comparisons: $g_{1,k}(x_i, x_j)$, $g_{2,k}(x_i, x_j)$ or $g_{3,k}(x_i, x_j)$ ($k = 1, \dots, N$; $N \geq 1$; $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$), under the assumption that difference $T_f(x_i, x_j) - g_{fk}(x_i, x_j)$ is a random error; moreover, the number of subsets n is assumed unknown. The following assumptions about comparison errors $T_f(\cdot) - g_{fk}(\cdot)$ ($f = 1, 2, 3$) are made in the paper:

A1. In the case of binary approach it is assumed that probability of incorrect comparison is lower than a correct one:

$$P(T_1(x_i, x_j) = g_{1,k}(x_i, x_j)) > P(T_1(x_i, x_j) \neq g_{1,k}(x_i, x_j)).$$

A2. In the case of multivalent approach it is assumed that: the probability function of a comparison error $P(T_f(x_i, x_j) - g_{fk}(x_i, x_j) = l)$ ($f = 2, 3$; $-(m-1) \leq l \leq m-1$) is unimodal with mode and median equal zero ($l = 0$).

A3. Multiple comparisons of the same pair ($N > 1$) are stochastically independent:

$$\begin{aligned} P((T_f(x_i, x_j) = g_{fk}(x_i, x_j)) \cap (T_f(x_r, x_s) = g_{fl}(x_r, x_s))) = \\ P((T_f(x_i, x_j) = g_{fk}(x_i, x_j))P(T_f(x_r, x_s) = g_{fl}(x_r, x_s)) \\ (l \neq k; j \neq i; s \neq r). \end{aligned}$$

A4. The comparison errors $T_2(x_i, x_j) - g_{2k}(x_i, x_j)$ and $T_3(x_i, x_j) - g_{3k}(x_i, x_j)$ ($1 \leq k \leq N$; $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$) are uncorrelated.

The assumption A1 is a special case of assumption A2, i.e. valid for a zero-one random variable. The assumptions A1–A4 can be verified with the use of statistical tests. The assumption A4 indicates that the comparisons $g_{2,k}(x_i, x_j)$ and $g_{3,k}(x_i, x_j)$ contain not the same information.

Each comparison $g_{fk}(x_i, x_j)$ evaluates the value $T_f(x_i, x_j)$; whole set of comparisons $g_{fk}(x_i, x_j)$ ($k = 1, \dots, N$; $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$) can be used as the basis for estimation of an unknown relation form. The estimators considered in the paper are based on the idea of minimization of differences between comparisons and the relation form. Two forms of differences are used as a basis for construction of estimators:

$$\sum_{\langle i, j \rangle \in R_m} \sum_{k=1}^N |t_f(x_i, x_j) - g_{fk}(x_i, x_j)|$$

and
$$\sum_{\langle i, j \rangle \in R_m} |t_f(x_i, x_j) - g_f^{(me)}(x_i, x_j)| \quad (f = 1, 2, 3), \quad (5)$$

where:

R_m – the set $R_m = \{\langle i, j \rangle \mid 1 \leq i, j \leq m; j > i\}$,

$t_f(x_i, x_j)$ – the function characterizing any tolerance relation χ_1, \dots, χ_r ,

$g_f^{(me)}(x_i, x_j)$ – the median in the set of comparisons $\{g_{f,1}(x_i, x_j), \dots, g_{fN}(x_i, x_j)\}$,

N – odd number.

It is clear that minimization of the first sum needs more computations than the second one, but can provide more efficient estimates.

4 Binary Approach

Minimization of the sums (5), in the case of binary comparisons $g_{1k}(x_i, x_j)$ ($k = 1, \dots, N$), indicate the estimators $\hat{\chi}_1^{(1,s)}, \dots, \hat{\chi}_n^{(1,s)}$ and $\hat{\chi}_1^{(1,me)}, \dots, \hat{\chi}_n^{(1,me)}$, obtained as optimal solutions of the tasks – respectively:

$$\min_{\chi_1^{(t)}, \dots, \chi_r^{(t)}} \left\{ \sum_{\langle i, j \rangle \in R_m} \sum_{k=1}^N |t_1^{(t)}(x_i, x_j) - g_{1k}(x_i, x_j)| \right\}, \quad (6)$$

$$\min_{\chi_1^{(t)}, \dots, \chi_r^{(t)}} \left\{ \sum_{\langle i, j \rangle \in R_m} |t_1^{(t)}(x_i, x_j) - g_1^{(me)}(x_i, x_j)| \right\}, \quad (7)$$

where:

F_X – the feasible set comprising all tolerance relations in the set \mathbf{X} ,

$\chi_1^{(t)}, \dots, \chi_r^{(t)}$ – t -th element of the set F_X ,

$t_1^{(t)}(\cdot)$ – the function describing the relation $\chi_1^{(t)}, \dots, \chi_r^{(t)}$.

Minimal values of the criterion functions (6)–(7) equal zero, the number of optimal solutions of each task may exceed one.

Under the assumptions (A1)–(A3), the estimates $\hat{\chi}_1^{(1,s)}, \dots, \hat{\chi}_{\hat{n}}^{(1,s)}$ and $\hat{\chi}_1^{(1,me)}, \dots, \hat{\chi}_{\hat{n}}^{(1,me)}$ have statistical properties resulting from the following facts (see [7]):

1⁰. Expected values of the random variables $W_{1N}^{(s)*}, W_{1N}^{(me)*}$ and $\tilde{W}_{1N}^{(s)}, \tilde{W}_{1N}^{(me)}$:

$$W_{1N}^{(s)*} = \sum_{\langle i,j \rangle \in R_m} \sum_{k=1}^N |T_1(x_i, x_j) - g_{1k}(x_i, x_j)|, \quad (8)$$

$$W_{1N}^{(me)*} = \sum_{\langle i,j \rangle \in R_m} |T_1(x_i, x_j) - g_1^{(me)}(x_i, x_j)|, \quad (9)$$

$$\tilde{W}_{1N}^{(s)} = \sum_{\langle i,j \rangle \in R_m} \sum_{k=1}^N |\tilde{t}_1(x_i, x_j) - g_{1k}(x_i, x_j)|, \quad (10)$$

$$\tilde{W}_{1N}^{(me)} = \sum_{\langle i,j \rangle \in R_m} |\tilde{t}_1(x_i, x_j) - g_1^{(me)}(x_i, x_j)|, \quad (11)$$

corresponding – respectively – to the actual relation $\chi_1^*, \dots, \chi_n^*$ and any relation $\tilde{\chi}_1, \dots, \tilde{\chi}_n$ different than actual, satisfy the inequalities:

$$E(W_{1N}^{(s)*} - \tilde{W}_{1N}^{(s)}) < 0, \quad (12)$$

$$E(W_{1N}^{(me)*} - \tilde{W}_{1N}^{(me)}) < 0, \quad (13)$$

$$P(W_{1N}^{(s)*} < \tilde{W}_{1N}^{(s)}) \geq 1 - \exp\{-2N(\frac{1}{2} - \delta)^2\}, \quad (14)$$

$$P(W_{1N}^{(me)*} < \tilde{W}_{1N}^{(me)}) \geq 1 - 2 \exp\{-2N(\frac{1}{2} - \delta)^2\}. \quad (15)$$

The inequalities (14), (15) are based on Hoeffding's inequalities (4); their right-hand side requires only the value of δ or its evaluation.

The inequalities (12) and (13) show that the values $E(W_{1N}^{(s)*}), E(W_{1N}^{(me)*})$, corresponding to actual relation $\chi_1^*, \dots, \chi_n^*$, are lower than the values – respectively – $E(\tilde{W}_{1N}^{(s)}), E(\tilde{W}_{1N}^{(me)})$, corresponding to any other relation $\tilde{\chi}_1, \dots, \tilde{\chi}_n$. Moreover the inequalities (14), (15) guarantee convergence to zero the probabilities of the events $\{W_{1N}^{(s)*} < \tilde{W}_{1N}^{(s)}\}, \{W_{1N}^{(me)*} < \tilde{W}_{1N}^{(me)}\}$, for $N \rightarrow \infty$; the speed of convergence is determined by exponential term: $\exp\{-2N(\frac{1}{2} - \delta)^2\}$. The inequalities (12)–(15) are the basis for estimators (6), (7), minimizing differences between estimates and comparisons $g_{1,k}(x_i, x_j)$ or $g_1^{(me)}(x_i, x_j)$ ($k = 1, \dots, N; \langle i, j \rangle \in R_m$). It is clear that the task (6) is easier to solve, than the task (7). However, simulation experiments show that the estimator (6) is more precise. The expected value of an estimation error:

$$\hat{\Delta}_{1,N}^{(s)} = \sum_{\langle i,j \rangle \in R_m} |T_1(x_i, x_j) - \hat{T}_{1,N}^{(s)}(x_i, x_j)|,$$

where:

$\hat{T}_{1,N}^{(s)}(\cdot)$ – an estimate of the function $T_1(\cdot)$, resulting from $\hat{\chi}_1^{(1,s)}, \dots, \hat{\chi}_{\hat{n}}^{(1,s)}$, is lower than the expected value of an error:

$$\hat{\Delta}_{1,N}^{(me)} = \sum_{\langle i,j \rangle \in R_m} \left| T_1(x_i, x_j) - \hat{T}_{1,N}^{(me)}(x_i, x_j) \right|$$

where:

$\hat{T}_{1,N}^{(me)}(\cdot)$ – an estimate of the function $T_1(\cdot)$, resulting from $\hat{\chi}_1^{(1,me)}, \dots, \hat{\chi}_{\hat{n}}^{(1,me)}$.

Moreover, the number of estimates with the same minimal value of the criterion function is greater in the case of the median estimator.

The properties of the estimators (6), (7) are similar to properties of estimators of the preference relation discussed in [5].

5 Multivalent Approach

Minimization of the sums (5), in the case of multivalent comparisons, indicate the estimators $\hat{\chi}_1^{(f,s)}, \dots, \hat{\chi}_{\hat{n}}^{(f,s)}$ and $\hat{\chi}_1^{(f,me)}, \dots, \hat{\chi}_{\hat{n}}^{(f,me)}$ ($f = 2, 3$), obtained as optimal solutions of the tasks – respectively:

$$\min_{\chi_1^{(t)}, \dots, \chi_r^{(t)}} \left\{ \sum_{\langle i,j \rangle \in R_m} \sum_{k=1}^N \left| t_f^{(t)}(x_i, x_j) - g_{fk}(x_i, x_j) \right| \right\}, \quad (f = 2, 3) \quad (16)$$

$$\min_{\chi_1^{(t)}, \dots, \chi_r^{(t)}} \left\{ \sum_{\langle i,j \rangle \in R_m} \left| t_f^{(t)}(x_i, x_j) - g_f^{(me)}(x_i, x_j) \right| \right\}, \quad (f = 2, 3) \quad (17)$$

where:

$F_X, \chi_1^{(t)}, \dots, \chi_r^{(t)}$ – the same as in formulas (6), (7),

$t_f^{(t)}(\cdot)$ ($f = 2, 3$) – the functions describing the relation $\chi_1^{(t)}, \dots, \chi_r^{(t)}$.

The use of two types of comparisons: $g_{2,k}(x_i, x_j)$ and $g_{3,k}(x_i, x_j)$ together indicate the estimators $\hat{\chi}_1^{(23,s)}, \dots, \hat{\chi}_{\hat{n}}^{(23,s)}$ and $\hat{\chi}_1^{(23,me)}, \dots, \hat{\chi}_{\hat{n}}^{(23,me)}$ resulting from the tasks – respectively:

$$\begin{aligned} & \min_{\chi_1^{(t)}, \dots, \chi_r^{(t)}} \left\{ \sum_{\langle i,j \rangle \in R_m} \sum_{k=1}^N \left| t_2^{(t)}(x_i, x_j) - g_{2k}(x_i, x_j) \right| \right. \\ & \quad \left. + \sum_{\langle i,j \rangle \in R_m} \sum_{k=1}^N \left| t_3^{(t)}(x_i, x_j) - g_{3k}(x_i, x_j) \right| \right\}, \quad (18) \\ & \min_{\chi_1^{(t)}, \dots, \chi_r^{(t)}} \left\{ \sum_{\langle i,j \rangle \in R_m} \left| t_2^{(t)}(x_i, x_j) - g_2^{(me)}(x_i, x_j) \right| \right. \\ & \quad \left. + \sum_{\langle i,j \rangle \in R_m} \left| t_3^{(t)}(x_i, x_j) - g_3^{(me)}(x_i, x_j) \right| \right\}. \quad (19) \end{aligned}$$

The minimal value of the functions (16)-(18) equals zero, the number of solutions of each task may exceed one.

Under the assumptions (A2)–(A4), the estimates $\hat{\chi}_1^{(f,s)}, \dots, \hat{\chi}_{\hat{n}}^{(f,s)}$ and $\hat{\chi}_1^{(f,me)}, \dots, \hat{\chi}_{\hat{n}}^{(f,me)}$ have statistical properties resulting from the following facts (see [7]):

2⁰. The expected values of the random variables $W_{fN}^{(s)*}$, $W_{fN}^{(me)*}$, $\tilde{W}_{fN}^{(s)}$, $\tilde{W}_{fN}^{(me)}$ ($f = 2, 3$):

$$W_{fN}^{(s)*} = \sum_{\langle i, j \rangle \in R_m} \sum_{k=1}^N |T_f(x_i, x_j) - g_{fk}(x_i, x_j)|, \quad (20)$$

$$W_{fN}^{(me)*} = \sum_{\langle i, j \rangle \in R_m} |T_f(x_i, x_j) - g_f^{(me)}(x_i, x_j)|, \quad (21)$$

$$\tilde{W}_{fN}^{(s)} = \sum_{\langle i, j \rangle \in R_m} \sum_{k=1}^N |\tilde{t}_f(x_i, x_j) - g_{fk}(x_i, x_j)|, \quad (22)$$

$$\tilde{W}_{fN}^{(me)} = \sum_{\langle i, j \rangle \in R_m} |\tilde{t}_f(x_i, x_j) - g_f^{(me)}(x_i, x_j)|, \quad (23)$$

corresponding – respectively – to the actual relation $\chi_1^*, \dots, \chi_n^*$ and any relation $\tilde{\chi}_1, \dots, \tilde{\chi}_n$ different than actual, satisfy the inequalities:

$$E(W_{fN}^{(s)*} - \tilde{W}_{fN}^{(s)}) < 0, \quad (24)$$

$$E(W_{fN}^{(me)*} - \tilde{W}_{fN}^{(me)}) < 0, \quad (25)$$

$$P(W_{fN}^{(s)*} < \tilde{W}_{fN}^{(s)}) \geq \geq 1 - \exp \left\{ - \frac{N(\sum_{T_f(\cdot) \neq \tilde{t}_f(\cdot)} E(|T_f(\cdot) - g_{f,1}(\cdot)| - |\tilde{t}_f(\cdot) - g_{f,1}(\cdot)|))^2}{2\vartheta^2(m-1)^2} \right\} \quad (26)$$

$$P(W_{fN}^{(me)*} < \tilde{W}_{fN}^{(me)}) \geq \geq 1 + \frac{1}{v(m-1)} E \left(\sum_{T_f(\cdot) \neq \tilde{t}_f(\cdot)} |T_f(\cdot) - g_f^{(me)}(\cdot)| - |\tilde{t}_f(\cdot) - g_f^{(me)}(\cdot)| \right), \quad (27)$$

where:

$T_f \neq \tilde{t}_f(\cdot)$ – the set $\{\langle i, j \rangle \mid (x_i, x_j) \in \mathbf{X} \times \mathbf{X}; j > i; T_f(x_i, x_j) \neq \tilde{t}_f(x_i, x_j)\}$,

ϑ – the number of elements of the set $\{\langle i, j \rangle \mid (x_i, x_j) \in \mathbf{X} \times \mathbf{X}; j > i; T_f(x_i, x_j) \neq \tilde{t}_f(x_i, x_j)\}$.

The interpretation of the inequalities (24)-(26) is similar to interpretation of the inequalities (12)-(15). The properties of the right-hand side of the inequality (27) are similar to those obtained for the preference relation ([8]).

Similar results can be formulated for the variables: $W_{2,N}^{(s)*} + W_{3,N}^{(s)*}$, $W_{2,N}^{(me)*} + W_{3,N}^{(me)*}$, $\tilde{W}_{2,N}^{(s)} + \tilde{W}_{3,N}^{(s)}$, $\tilde{W}_{2,N}^{(me)} + \tilde{W}_{3,N}^{(me)}$, corresponding to the estimators (18), (19).

The right-hand sides of the inequalities (26), (27) require distributions of comparison errors and can be determined for fixed relation $\tilde{\chi}_1, \dots, \tilde{\chi}_n$; in the case of the median estimator it is necessary to determine the distributions of medians (see [1], section 2.4, [8]). Typically the distributions are not known; in such case they can be estimated (for appropriate N – at least several) or replaced by some approximation

(an example of approximation is presented in [7]). The distributions of errors of estimators (16)-(19) can be obtained with the use of simulation approach.

6 Summary and Conclusions

The paper presents the estimators of the tolerance relation based on pairwise comparisons with random errors developed by the author. Important feature of the estimators are weak assumptions about comparison errors. Especially, expected values of errors may be different than zero and the distributions of errors can be unknown. The assumptions allow combining comparisons obtained from different sources: statistical tests, experts' opinions, neural networks or other decision procedures. The estimates, obtained on the basis of discrete programming tasks, are not influenced by any arbitrary parameters or functions. They have good statistical properties. The approach can be considered, in some cases, as an alternative to fuzzy formalization.

Some properties of the estimators are difficult to determine in analytic way, e.g. probability function of estimation errors; they can be obtained with the use of simulation approach.

Acknowledgement. The work has been supported by the grant No N N111434937 of the Polish Ministry of Science and Higher Education.

References

- [1] David, H.A.: Order Statistics. J. Wiley, New York (1970)
- [2] David, H.A.: The Method of Paired Comparisons, 2nd edn. Ch. Griffin, London (1988)
- [3] Gordon, A.D.: Classification, 2nd edn. Chapman & Hall/CRC, Boca Raton (1999)
- [4] Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13–30 (1963)
- [5] Klukowski, L.: Some probabilistic properties of the nearest adjoining order method and its extensions. Annals of Operations Research 51, 241–261 (1994)
- [6] Klukowski, L.: Tests for relation type – equivalence or tolerance – in finite set of elements. Control and Cybernetics 35, 369–384 (2006)
- [7] Klukowski, L.: Estimation of tolerance relation the basis of multiple pairwise comparisons with random errors. Control and Cybernetics 36, 443–466 (2007a)
- [8] Klukowski, L.: Estimation of the Preference Relation on the Basis of Medians from Pairwise Comparisons in the Form of Difference of Ranks. In: Kurzynski, M., et al. (eds.) Computer Recognition Systems 2. ASC, vol. 45, pp. 232–241. Springer, Heidelberg (2007b)
- [9] Klukowski, L.: Determination of Tolerance Relation - alternative Approach to Intuitionistic and Fuzzy Sets. In: Atanassov, K.T., et al. (eds.) Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Applications, vol. 2, pp. 85–94. Ac. Publishing House EXIT, IBS PAN, Warsaw (2008)
- [10] Slater, P.: Inconsistencies in a schedule of paired comparisons. Biometrika 48, 303–312 (1961)

Radial Basis Function Kernel Optimization for Pattern Classification

Paweł Chudzian

Abstract. The main advantage of the kernel methods is the possibility of using linear models in a nonlinear subspace by an implicit transformation of patterns to a high-dimensional feature space without computing their images directly. An appropriately constructed kernel results in a model that fits well to the structure underlying data and doesn't over-fit to the sample. Recent state-of-the-art kernel evaluation measures are examined in this paper and their application in kernel optimization is verified. Alternative evaluation measures that outperform presented methods are proposed. Optimization leveraging these measures results in parameters corresponding to the classifiers that achieve minimal error rate for *RBF* kernel.

1 Introduction

Kernel-based learning methods have gained much attention in recent years, mainly due to high performance they provide in a variety of machine learning tasks [1]. Incorporating so called „kernel trick” [1] amounts to an implicit transformation of patterns to a high-dimensional (possibly even infinite) feature space without computing their images directly. A function $\phi : X \mapsto H$ mapping from an input space X to a feature space H is incorporated in a kernel via dot product:

$$k(\mathbf{v}_i, \mathbf{v}_j) = \langle \phi(\mathbf{v}_i), \phi(\mathbf{v}_j) \rangle. \quad (1)$$

An appropriately constructed kernel results in a model that fits well to the structure underlying data and doesn't over-fit to the data sample therefore the choice of a kernel function and its parameters is of crucial importance.

For a large number of learning tasks the choice of the off-the-shelf kernel is often sufficient [4]. Frequent choice is the Radial Basis Function (RBF) kernel:

Paweł Chudzian

Faculty of Electronics and Information Technology, Warsaw University of Technology,
Nowowiejska 15/19, 00-665 Warsaw, Poland

e-mail: P.Chudzian@elka.pw.edu.pl

$$k(\mathbf{v}_i, \mathbf{v}_j) = \exp\left(-\frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{2\sigma^2}\right). \quad (2)$$

The main challenge is to select optimal value of the σ parameter for a given classification problem. Methodology that involves a grid search over a space of kernel parameters along with creation and evaluation of a classifier model suffers from computational difficulties. This is especially true when the algorithm of choice is Support Vector Machine known for its relatively high complexity [2]. Moreover, grid search – due to limited resolution of the grid – doesn't provide a way of finding optimal parameters.

Recent work on kernel evaluation measures, e.g. [5, 9, 12, 13] provide a basis for the optimization of kernel parameters, however experiments show that none of the recent state-of-the-art measures result in a classifiers of minimal error rate. In this paper alternative kernel evaluation measures that outperform standard methods are proposed along with the methodology which enables finding appropriate kernel parameters without a need to build a classification model. RBF kernel optimization leveraging proposed evaluation measures results in parameters corresponding to the classifier that achieves the minimal error rate while the complexity of the model is controlled. In addition presented measures allow to apply convex optimization methods to effectively tune kernel parameters. Results obtained on benchmark data sets from UCI repository show that kernel evaluation measures proposed in this paper are better suited for RBF kernel optimization than methods presented in the literature. The methodology is additionally verified on a real-world problem of classification of objects detected with an impulse radar.

The rest of the paper is organized as follows. Section 2 presents recent work on kernel evaluation measures. In section 3 alternative kernel evaluation measures are presented. The methodology for finding kernel parameters and experimental results are shown in section 4. Last section presents conclusions and future work.

2 Related Work

Recent years have brought the problem of kernel parameters selection to the broader attention. Beside the grid search methodology there are two main approaches to solve the task. First consist in optimization of a kernel matrix containing kernel function evaluations for all pairs of n patterns in a given data sample:

$$K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j), \quad i, j = 1, \dots, n. \quad (3)$$

Given a set of class labels for a data sample one can indirectly optimize kernel operating on a kernel matrix. This approach is usually used for transductive inference [5] i.e. when test cases are available beforehand. In second approach kernel function parameters are optimized directly which enables inductive learning as well as overcomes the limitations of the grid search.

Regardless of the methodology it is necessary to provide an objective function for the optimization that would result in a classifier of high accuracy and wouldn't

require complex calculations. The set of standard kernel evaluation measures satisfying these conditions consist of Kernel Target Alignment and Fisher-based measures.

2.1 Kernel Target Alignment

Though introduced almost a decade ago Kernel Target Alignment (KTA) [5, 7] is still most widely used kernel evaluation measure, mainly due to its effectiveness and low complexity [8, 10, 12]. It computes how good a kernel matrix K aligns to a class matrix Y . The latter take values representing maximal and minimal similarity of patterns of the same and different classes respectively. In context of pattern classification class matrix Y represents the optimal kernel matrix. Alignment evaluation is based on the cosine calculation between two-dimensional vectors represented by matrices. Kernel Target Alignment is a special case of this measure where kernel and class matrices are involved:

$$A(K, Y) = \frac{\langle K, Y \rangle_F}{\sqrt{\langle K, K \rangle_F \langle Y, Y \rangle_F}}, \quad (4)$$

where $\langle \cdot, \cdot \rangle_F$ denotes Frobenius product defined as $\langle A, B \rangle_F = \sum_{i,j=1}^n A_{ij} B_{ij}$.

Kernel Target Alignment takes values in $[-1; 1]$ range. The higher the KTA value the more aligned both matrices hence the kernel provides high class separability and good classification results. Though high KTA values amount to the kernel well aligned to the class distribution the opposite is not always true. Analysis of KTA presented in [9] shows that the kernel matrix yielding perfect class separability may achieve low KTA value.

2.2 Fisher-Based Kernel Evaluation Measure

Kernel evaluation measures presented in [12, 13] are based on Fisher's linear discrimination rule. Optimization with such objective functions corresponds to the maximization of class separability in a feature space H :

$$J = \frac{\text{tr}(S_B)}{\text{tr}(S_W)}, \quad (5)$$

where S_B and S_W represent between-class and within-class scatter matrices respectively. Results with Fisher-based evaluation measure obtained in [13] using gradient descent optimization show that the optimized RBF kernel yields significant improvement of the classification with linear classifiers. On the other hand methodology used there for the model evaluation make the results biased and hard to reproduce. For each data set authors perform normalization and then random splitting into three equal and disjoint parts, one for each of kernel optimization, model training and model testing. Reported results are averaged over 20 runs. There are two issues

with this design. First of all normalization should be carried out on a training set and then obtained parameters should be applied to the data sample held out for testing. This mimics real-world situation when there is no information about new data before it arrives. The second issue regards splitting schema – in case of the cross-validation each pattern is used exactly once for model evaluation. In the schema adopted in [13] there is no guarantee that all patterns are used during training and model evaluation.

2.3 Feature Space-Based Kernel Matrix Evaluation Measure

Modification of the Fisher-based approach that replaces between-class scatter matrix with distance between class centers was proposed in [9]. Feature Space-based Kernel Matrix Evaluation Measure (FSM) is defined as the ratio of the total within-class variance v_w in the direction between the class centers to the distance between these centers:

$$FSM(K, y) = \frac{v_w}{\|\bar{\Phi}^{(1)}(\mathbf{x}) - \bar{\Phi}^{(2)}(\mathbf{x})\|}, \quad (6)$$

where $\bar{\Phi}^{(i)}(\mathbf{x})$ denotes i -th class center in a feature space H . Finally, the error rate $FSMerr = \frac{FSM^2}{1+FSM^2}$ is used to keep the values of the measure in $[0; 1]$ range.

Fisher's discrimination rule is known to be vulnerable to heteroscedastic distributions. This also affects measures that are based on that rule. Figure 1 shows artificial data generated from different distributions along with the $FSMerr$ values for these data sets. $FSMerr$ increases from left to right on the figure while it is clear that the rightmost case is linearly separable and should receive $FSMerr$ value comparable to the leftmost one. It is clear then that $FSMerr$ may substantially misjudge particular kernels.

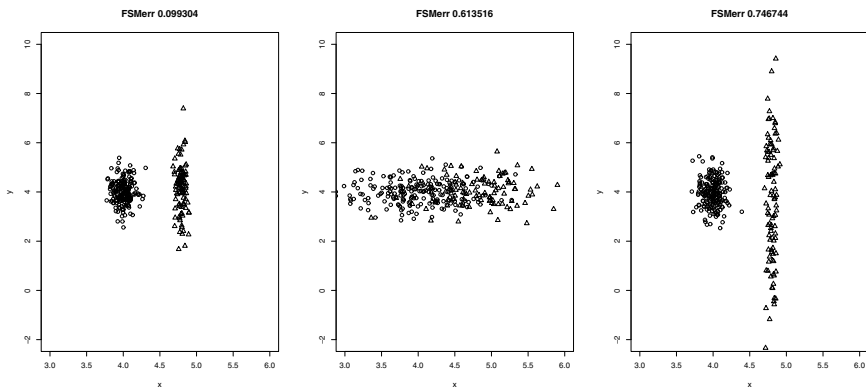


Fig. 1 $FSMerr$ values for the artificial data generated from different distribution.

Despite possessing evident drawbacks presented kernel evaluation measures are still widely used in the process of kernel optimization for pattern classification. Their complexity is relatively low as there is only one pass through the kernel matrix necessary to calculate each value. Results shown in [5, 7, 9, 12, 13] suggest relatively high improvement of the classification accuracy when measures are used. However methodologies – both in the context of data sets used for experiments and classification algorithms involved – presented in each work differs significantly making the comparison impossible.

3 Kernel Evaluation

Alternative kernel evaluation measures presented in this section strive to overcome weaknesses related to the previously introduced methods and exhibit low complexity at the same time.

3.1 Measure α – Correlation Based

Measure α is based on the Pearson correlation coefficient evaluated between normalized kernel values¹ and the class distances between respective patterns. The Class distance equals 1 if patterns belong to the same class and 0 otherwise. The basic idea behind this measure is that the farther the patterns lie from each other (normalized kernel value approaches 0) the bigger the distance between their class labels is expected. Measure α takes values in $[-1; 1]$ range with higher values corresponding to the kernel better aligned to the class distances between patterns.

3.2 Measure β – Means Equality Test

Measure β is based on the t -test of equality of means of two groups of kernel values. One group consist of kernel values evaluated on patterns that belong to the same class. Kernel values defined on patterns from opposite classes belong to the other group. The idea is that for a separable case patterns of a particular category should lie close to each other and simultaneously far from patterns of the other class. In such case the mean within-class kernel value should greatly differ from the mean between-class kernel value. When a test is applied to such data the value of the statistic is high. In general, the higher the separability of classes in a feature space H is observed the higher the value of the statistic and hence the β measure.

3.3 Measure γ – KTA Modification

KTA values strongly depend on the class cardinalities – in case of strong heterogeneity of the class distribution sensitivity of KTA drops drastically. Measure γ provides normalization that reduces impact of intra-class factor:

¹ Normalized kernel is defined as $\tilde{k}(\mathbf{x}_1, \mathbf{x}_2) = k(\mathbf{x}_1, \mathbf{x}_2) / \sqrt{k(\mathbf{x}_1, \mathbf{x}_1)k(\mathbf{x}_2, \mathbf{x}_2)}$.

$$A_\gamma(K, y) = \frac{\frac{\sum_{i,j=1}^{n_1} K_{ij} + \sum_{i,j=n_1+1}^{n_2} K_{ij}}{n_1^2 + n_2^2} - \frac{\sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_2} K_{ij}}{n_1 n_2}}{\sum_{i,j=1}^n K_{ij}^2}. \quad (7)$$

When the class cardinalities are equal $A_\gamma(K, y) = A(K, y)$. When the class distribution heterogeneity increases the left-hand side of the nominator is being decreased faster than the right-hand side and keeps the sensitivity of $A_\gamma(K, y)$ on a reasonable level.

4 Experiments

Experiments were performed on 5 data sets from UCI repository [6]. *Wisconsin Breast Cancer* (699 instances; 10 attributes), *Pima Indians Diabetes* (768; 9) and *Johns Hopkins University Ionosphere* (351; 34) are binary problems whereas *Balance Scale* (625; 4) and *Vehicle* (846; 18) consist of observation from 3 and 4 categories respectively. Data sets were chosen to differ in size (number of instances), complexity (number of attributes) and difficulty (in context of the classification task). The performance of proposed methodology and alternative kernel evaluation measures was also examined on a real-world problem of classification of objects detected with an impulse radar [7] (2036; 64).

Source code for the experiments was written in R and C++ languages with use of the *LIBSVM* library [3].

4.1 Methodology

For the verification of presented kernel evaluation measures the following methodology was applied. For each binary problem (multi-class data sets were decomposed into series of binary problems in the one-vs-one scheme) optimization was performed using the quasi-Newton *BFGS* algorithm with kernel evaluation measures as objective functions. Optimization was run several times with randomly picked starting point showing convergence to the same optimum each time a particular objective function was chosen. Obtained kernel parameters were subsequently used with the *SVM* algorithm to build the model and the classification error was determined. The ν -*SVM* algorithm was used with $\nu = 4$ in order to control the complexity of the model. Additional experiments show that for almost all considered binary problems RBF kernel parameter value for which classification error minimum occurs does not change with varying ν .

In order to assert the quality of the results and make them easily reproducible 10-fold cross-validation was applied 5 times and the results were averaged. That said every optimization-classification step along with the model evaluation was repeated 50 times on a different sample each time. This methodology enabled direct comparison of the presented standard and alternative kernel evaluation measures.

² Data set available courtesy of prof. Marek Nałęcz of Institute of Electronic Systems, Faculty of Electronics and Information Technologies, Warsaw University of Technology.

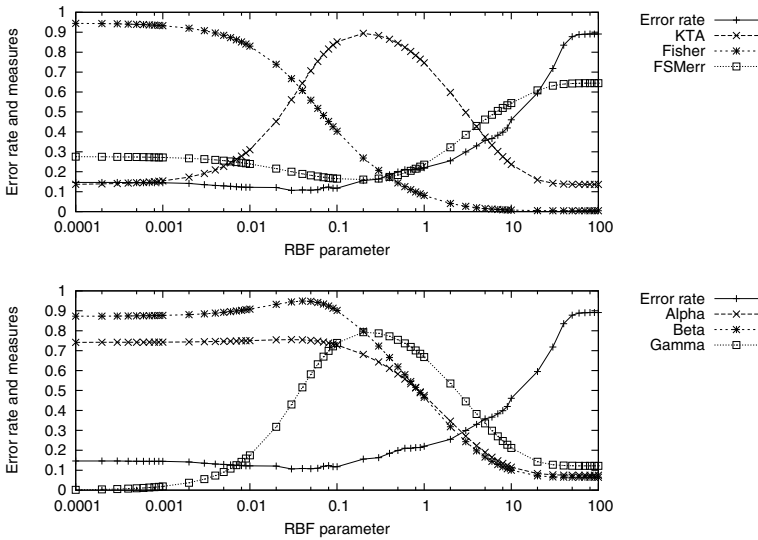


Fig. 2 The error rate and evaluation measures with different RBF parameter values.

4.2 Results

Figure 2 presents the error rate and values of evaluation measures for RBF kernel parameter running from 10^{-4} to 100. For the presentation purposes graphs were scaled to fit into $[0; 1]$ range. Only for α and β measures maxima match minimum of the error rate for the SVM classifier. The worst results are achieved with Fisher-based measure. Such good extrema matching is not always the case with other data sets, however further experimental results show that using alternative kernel evaluation measures always give better or at least comparable results.

Table 1 presents mean error rates along with variance of error obtained with the SVM classifier using the optimized kernel. Best results are marked in bold. Models built using the RBF kernel with parameter optimized using *KTA* achieve small classification error comparable to the best results obtained with the SVM classifier reported in the literature. On the other hand both Fisher-based measure and *FSMerr* fail to give good RBF kernel parameter except a few binary tasks.

The results for models built using alternative kernel evaluation measures are presented in table 2. The lowest error amongst these measures is marked in bold and all results better than the lowest error from table 1 are additionally marked with italics. The best overall results are achieved for α and β measures. Former gives the best results in 7 out of total 12 binary problems (with exactly half of the results outperforming the best of the standard measures). The measure β also gives the lowest error rate in 7 tasks and outperforms standard measures in 9 out of 12 situations. Modification of *KTA* (γ measure) although best only for 2 data sets is still better than *KTA* and Fisher-based measures in 8 out of 12 comparisons.

Table 1 Error rates for the SVM built over *KTA* and Fisher-based measures.

	<i>KTA</i>		Fisher		<i>FSMerr</i>	
Breast Cancer	0.0351	(1e-06)	0.0430	(1.5e-05)	0.0354	(3.6e-06)
Pima Diabetes	0.3010	(4.2e-05)	0.3765	(1.7e-02)	0.3424	(3.6e-05)
Ionosphere	0.0541	(1.2e-05)	0.0547	(9.7e-06)	0.0541	(1.2e-05)
Balance-Scale (L – B)	0.0522	(2.6e-06)	0.5287	(2.8e-02)	0.1139	(3.1e-04)
Balance-Scale (L – R)	0.0229	(3.6e-06)	0.1350	(5.3e-03)	0.0302	(9.9e-06)
Balance-Scale (B – R)	0.0540	(2.8e-05)	0.4534	(1.9e-02)	0.1115	(5.2e-04)
Vehicle (Opel – Saab)	0.4526	(3.7e-04)	0.2913	(5.5e-04)	0.4932	(9.7e-06)
Vehicle (Opel – Bus)	0.0255	(5.4e-06)	0.0144	(9.1e-06)	0.5046	(2.7e-06)
Vehicle (Opel – Van)	0.0574	(2.8e-05)	0.2642	(1.6e-03)	0.4793	(1.1e-06)
Vehicle (Saab – Bus)	0.0202	(1e-06)	0.0211	(3.6e-06)	0.4979	(3.5e-05)
Vehicle (Saab – Van)	0.0389	(1.2e-05)	0.2687	(7.2e-04)	0.4745	(4.6e-06)
Vehicle (Van – Bus)	0.0163	(6.9e-06)	0.4201	(8.5e-03)	0.1563	(7.7e-03)

Table 2 Error rates for the SVM built over alternative measures.

	α		β		γ	
Breast Cancer	0.0286	(1.7e-06)	0.0298	(6.4e-07)	0.0363	(4.2e-07)
Pima Diabetes	0.3031	(6.2e-05)	0.3010	(1.1e-04)	0.2976	(4.7e-05)
Ionosphere	0.0484	(4e-06)	0.0541	(8.1e-06)	0.0541	(1.2e-05)
Balance-Scale (L – B)	0.3080	(1.3e-02)	0.0468	(3.2e-05)	0.0510	(5.8e-05)
Balance-Scale (L – R)	0.0138	(0)	0.0138	(0)	0.0218	(2.4e-06)
Balance-Scale (B – R)	0.3483	(1.7e-02)	0.0498	(4.5e-05)	0.0510	(2.8e-05)
Vehicle (Opel – Saab)	0.4046	(9.4e-05)	0.4102	(1.6e-04)	0.2755	(3.2e-04)
Vehicle (Opel – Bus)	0.0251	(6.4e-06)	0.0251	(3.7e-06)	0.1200	(8.1e-03)
Vehicle (Opel – Van)	0.0389	(1.1e-05)	0.0379	(2.8e-05)	0.0506	(1.8e-05)
Vehicle (Saab – Bus)	0.0206	(0)	0.0206	(0)	0.0211	(1e-06)
Vehicle (Saab – Van)	0.0331	(3e-05)	0.0365	(3.2e-05)	0.0408	(6.9e-05)
Vehicle (Van – Bus)	0.0163	(1.1e-06)	0.0153	(7.4e-06)	0.0177	(4.6e-06)

Table 3 Error rates for the SVM built over *KTA* and Fisher-based measures.

	<i>KTA</i>		Fisher		<i>FSMerr</i>	
Cessna – Jak 12	0.2618	(1.8e-05)	0.4681	(0.0029)	0.4931	(0)
Cessna – Mi-2	0.0332	(1e-06)	0.1151	(0.0043)	0.0286	(4.4e-07)
Cessna – Mi-8	0.0788	(1.1e-05)	0.2769	(0.015)	0.0778	(3.4e-06)
Cessna – Mig-29	0.0127	(0)	0.0776	(0.0013)	0.0198	(4.8e-07)
Jak 12 – Mi-2	0.0428	(1.1e-05)	0.1858	(0.0054)	0.0349	(1.9e-05)
Jak 12 – Mi-8	0.0758	(1.1e-05)	0.1616	(0.00017)	0.0768	(1e-05)
Jak 12 – Mig-29	0.0123	(1.3e-06)	0.0438	(0.00065)	0.0128	(0)
Mi-2 – Mi-8	0.0592	(5.4e-06)	0.2085	(0.00065)	0.1991	(0.0030)
Mi-2 – Mig-29	0.0283	(4e-07)	0.1041	(0.0030)	0.0400	(1.1e-06)
Mi-8 – Mig-29	0.0207	(1.6e-06)	0.2135	(0.0046)	0.0205	(1.7e-06)

The results for the radar data presented in table 3 show that using *KTA* for kernel optimization results in the best model in 8 out of 10 binary tasks. Models based on the kernel optimized with *FSMerr* give the smallest error rate in 4 problems. Fisher-rule based measure fails as an objective function raising models of radically worse accuracy with the classification error up to an order of magnitude higher than *KTA* and *FSMerr*.

Using alternative kernel evaluation measures for the radar data set (table 4) shows further improvement in the accuracy of models. In case of β measure results are better than the best of those achieved with the models built using standard measures

Table 4 Error rates for the SVM built over alternative measures.

	α	β	γ
Cessna – Jak 12	0.2158 (6.9e-05)	0.2144 (5.6e-05)	0.2653 (3.4e-05)
Cessna – Mi-2	0.0298 (4.4e-07)	0.0283 (1e-06)	0.0312 (1.9e-06)
Cessna – Mi-8	0.0798 (1.2e-05)	0.0783 (3.3e-06)	0.0820 (1.1e-05)
Cessna – Mig-29	0.0132 (4.8e-07)	0.0124 (3.2e-07)	0.0127 (0)
Jak 12 – Mi-2	0.0356 (5.6e-06)	0.0339 (1.9e-06)	0.0415 (2e-05)
Jak 12 – Mi-8	0.0738 (2.7e-06)	0.0703 (3.5e-06)	0.0771 (8.5e-06)
Jak 12 – Mig-29	0.0110 (1.3e-06)	0.0113 (1.1e-06)	0.0134 (1.3e-06)
Mi-2 – Mi-8	0.0601 (1.8e-05)	0.0612 (7.9e-06)	0.0581 (6.7e-06)
Mi-2 – Mig-29	0.0276 (9.4e-07)	0.0276 (9.4e-07)	0.0283 (4e-07)
Mi-8 – Mig-29	0.0200 (2.7e-07)	0.0205 (2.4e-06)	0.0212 (2.7e-07)

Table 5 Errors rate for the SVM – multi-class problem.

	<i>KTA</i>	Fisher	<i>FSMerr</i>
Balance-scale	0.050 (1.6e-05)	0.496 (9.8e-03)	0.097 (7.7e-05)
Vehicle	0.267 (1.2e-04)	0.461 (1.6e-03)	0.722 (2.7e-04)
Radar	0.184 (3.2e-05)	0.420 (6.3e-03)	0.288 (2.9e-04)
	α	β	γ
Balance-scale	0.378 (1.4e-02)	0.050 (1.6e-05)	0.055 (8.1e-05)
Vehicle	0.237 (1.2e-04)	0.260 (4.4e-04)	0.179 (6.4e-05)
Radar	0.159 (1.3e-05)	0.159 (9e-06)	0.173 (5.2e-06)

in 9 out of 10 binary tasks. Using α for optimization results in a decrease of the error rate in 5 tasks. Measure γ doesn't give much of the improvement over the *KTA* for radar data as the almost uniform class distribution makes both measures very similar.

In order to test the quality of the evaluation measures for multi-class tasks the optimal kernels for each decomposed binary tasks were used in the one-vs-one schema with voting to evaluate the error rate for the whole multi-class problem. The results are shown in table 5. *KTA* gives the best models among three standard measures – using Fisher-based method and *FSMerr* result in substantially worse models. Application of alternative measures results in improvement of the accuracy of classifiers over those obtained with the RBF kernel optimized using *KTA*. For *Balance-Scale* data set optimization with β measure results in the SVM classifier of the accuracy equal to the model built after optimization with *KTA*. For *Vehicle* and *Radar* data sets using alternative kernel evaluation measures results in the classifier with the error rate several percent smaller than those obtained after the optimization with *KTA*.

5 Conclusions and Future Work

In this paper novel approaches to the evaluation of kernels were proposed that outperform state-of-the-art kernel evaluation measures and enable efficient calculation. Kernel optimization leveraging suggested evaluation measures results in parameters corresponding to the classifiers that achieve substantially smaller error rates for the RBF kernel function than those built with kernels optimized using standard evaluation measures. In addition, presented measures allow to apply convex optimization methods which enables effective tuning of the kernel parameters.

Beside successful application to the RBF kernel initial experiments show promising results also with the polynomial kernel. Another step further is to examine if the absolute values of the evaluation measures correspond to the absolute values of error rates on different kernel functions. Experiments presented in [13] provide also the basis to the application of kernel evaluation measures in assessing the quality of kernel combinations. Natural step further is to investigate kernels combinations optimized with evaluation measures that result in models of high accuracy.

References

- [1] Aizerman, A., Braverman, E.M., Rozoner, L.I.: Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning. ARC 25, 821–837 (1964)
- [2] Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery* 2(2), 121–167 (1998)
- [3] Chang, C.C., Lin, C.J.: LIBSVM: a library for Support Vector Machines (2001), <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [4] Cristianini, N., Shawe-Taylor, J.: *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, Cambridge (2000)
- [5] Cristianini, N., Shawe-Taylor, J., Elisseeff, A., Kandola, J.: On Kernel-Target Alignment. In: *Advances in Neural Information Processing Systems*, vol. 14, pp. 367–373. MIT Press, Cambridge (2001)
- [6] Frank, A., Asuncion, A.: UCI Machine Learning Repository (2010), <http://archive.ics.uci.edu/ml>
- [7] Kandola, J., Shawe-Taylor, J., Cristianini, N.: On the Extensions of Kernel Alignment. Technical Report NC-TR-02-120, Neural Networks and Computational Learning Theory (2002)
- [8] Kandola, J., Shawe-Taylor, J., Cristianini, N.: Optimizing Kernel Alignment over Combinations of Kernels. Technical Report at Department of Computer Science, Royal Holloway, University of London, UK (2002)
- [9] Nguyen, C.H., Ho, T.B.: Kernel Matrix Evaluation. In: Veloso, M.M. (ed.) *IJCAI*, pp. 987–992 (2007)
- [10] Pothin, J.B., Richard, C.: Optimal Feature Representation For Kernel Machines Using Kernel-Target Alignment Criterion. In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing* (2007)
- [11] Schölkopf, B., Smola, A.J.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge (2001)
- [12] Wang, J., Lu, H., Plataniotis, K.N., Lu, J.: Gaussian Kernel Optimization for Pattern Classification. *Pattern Recognition* 42(7), 1237–1247 (2009)
- [13] Xiong, H., Swamy, M.N.S., Ahmad, M.O.: Optimizing the Kernel in the Empirical Feature Space. *IEEE Transactions on Neural Networks* 2(16), 460–474 (2005)

Evaluation of Reliability of a Decision-Making Process Based on Pattern Recognition

Artur Sierszeń and Łukasz Sturgulewski

Abstract. In recent years, all kinds of assistance to decision-making processes have become of great importance. In many areas of life, technology, and science, computation machines, i.e. computers, assist people in making decisions; in some particular cases, computers are even able to make these decisions themselves. There are, admittedly, many situations, i.e. in flight control, airport traffic, or biomedical data analysis, where even decisions made by the most efficient algorithm (or the most efficient machine) must be confirmed by a man. Nevertheless, in many disciplines, such as finance, psychology, medicine, or biology, mechanisms assisting the decision-making process facilitate work significantly. One of the key issues is the reliability of made decisions. However, this issue is often passed over due to great increase in the amount of data processed by those systems, what forces the designers to focus on the systems' efficiency.

1 Introduction

Pattern recognition consists in determining the affiliation of certain objects to defined classes; objects are understood very generally and are representations of items, phenomena, processes, or events. Each object described by a set of features is treated as a point in an n -dimensional feature space. A recognition algorithm may be treated as assigning the number of a class i to an object x , i.e. reproducing a feature space K in a set of class numbers. The selection of an appropriate representation K decides about the usefulness of a recognition method in question to a specific purpose. This explains a large number of different algorithms created in this discipline. The most frequently used pattern recognition applications include classification of documents, e.g. searching for documents in

Artur Sierszeń · Łukasz Sturgulewski
Technical University of Łódź,
Computer Engineering Department,
ul. Stefanowskiego 18/22, 90-924 Łódź
e-mail: asiersz, luk@kis.p.lodz.pl

the Internet, identification of people based on their faces, irises, or fingerprints, writing recognition, speech recognition, quality control in industry, classification of microscope images or X-ray images, and classification or other medical data. Each of these purposes has different characteristics; therefore, each of them requires different approach. For each of them, a method should be selected from a great number of pattern recognition methods, one which will make it possible to achieve the best possible results. There are two main factors deciding about the usefulness of a pattern recognition algorithm to a specific purpose, namely classification time and classification error. In an extensive literature on pattern recognition, there has been a lot of discussions on minimum-distance classifiers of various types and some related issues, such as selection of features or reduction of a reference set. Many methods for estimation of classification error have also been proposed. Evaluation of reliability was also discussed [2]; however, the authors have not managed to find any information on controlling reliability in the aspect of classification. The possibility of evaluating reliability of decisions depending on features of classified objects is an additional parameter steering between the error and the classification speed; in the author's research it is also one of the basic parameters. The algorithm will be used in the authors' implementation of network behavioural analysis system using the learning set and the decision rules based on distance functions. In systems of this type, efficiency and quality aspects are important as well the assessment of decision certainty. The aim of the research is to develop a mechanism for detecting irregularities using pattern recognition algorithms. [3]

2 Description of the Algorithm

The first stage of the algorithm's operation (Fig. 1) includes reading data from a configuration file which contains information on the name of the source file and the resulting file, the classification algorithm to be applied, and its parameters (for k -NN, these information include the k value, the percentage division into the learning set and the testing set, and the method to determine the level of trust). During the next stage, data from the source file is imported. The format of the file must be consistent with the commonly adopted standard [1].

During the first stage (Fig. 2), classes are selected which will be used in further computations (it may happen that some classes imported from the file are not significant). This information is passed to the k -NN function through the `_bTabUser` table of the `_uiIndexUser` size. The next parameter passed to the k -NN function is `_uiK`, which determines the number of nearest neighbours taken into account in the classification process. If the parameter's value is 0, the number of nearest neighbours will be determined automatically (from the range of $k = 1, 3, \dots, 21$) (Fig. 2).

During the next stage (Fig. 3), the level of classification quality is determined as well as the level of trust in the achieved results. Two methods for the evaluation

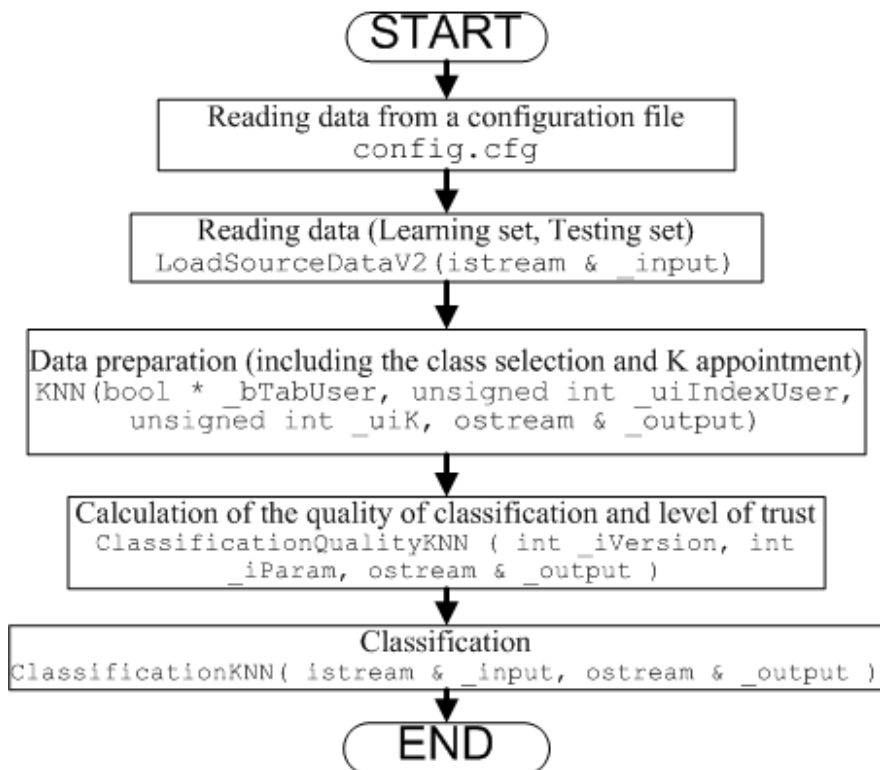


Fig. 1 General diagram of the program's operation.

of classification quality have been implemented, i.e. the resubstitution method and the allocation method. If the resubstitution method is selected, classification quality is evaluated for each object from the learning set. Then, the object is deleted from the set and classified with the k -NN method; finally, the result of classification is compared to the real class of the object. If the result is wrong, the *_dBad* counter (counting wrong decisions) is increased. In the allocation method, a testing set is allocated from the imported learning set (proportions for the division are determined by the user). Then, each object from the testing set is classified with the k -NN method; finally, the result of classification is compared to the real class of the object. If the result is wrong, the *_dBad* counter (counting wrong decisions) is increased. Additionally, the authors implemented the evaluation of the level of trust in the achieved classification result. It consists in determining what percentage of votes in the k -NN method was cast for the selected class (the *dTrust* variable).

Finally, the examined objects are classified with the k -NN method and the level of trust in the achieved results is evaluated (Fig. 4).

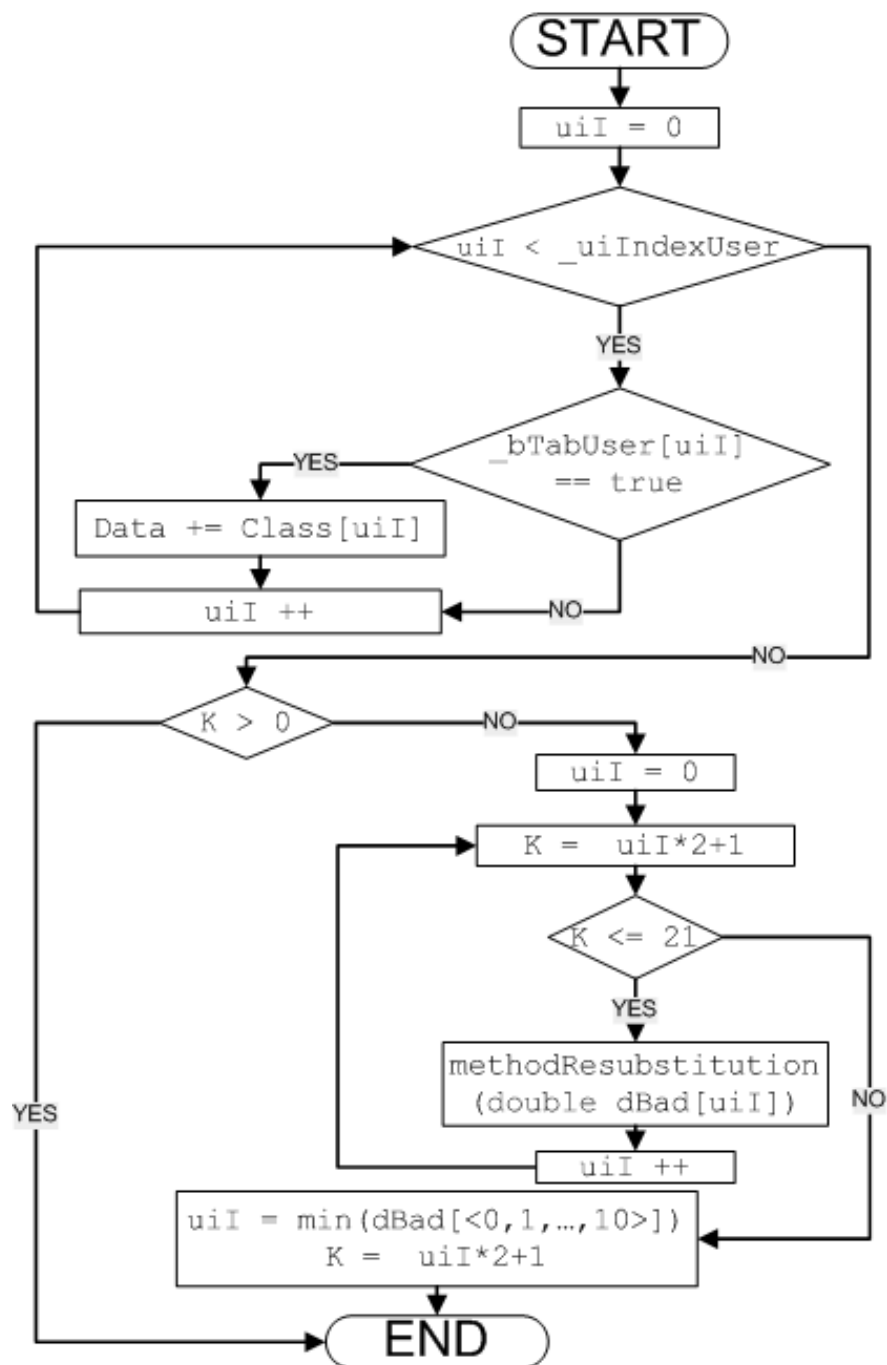


Fig. 2 The first stage of the algorithm's operation - selection of classes.

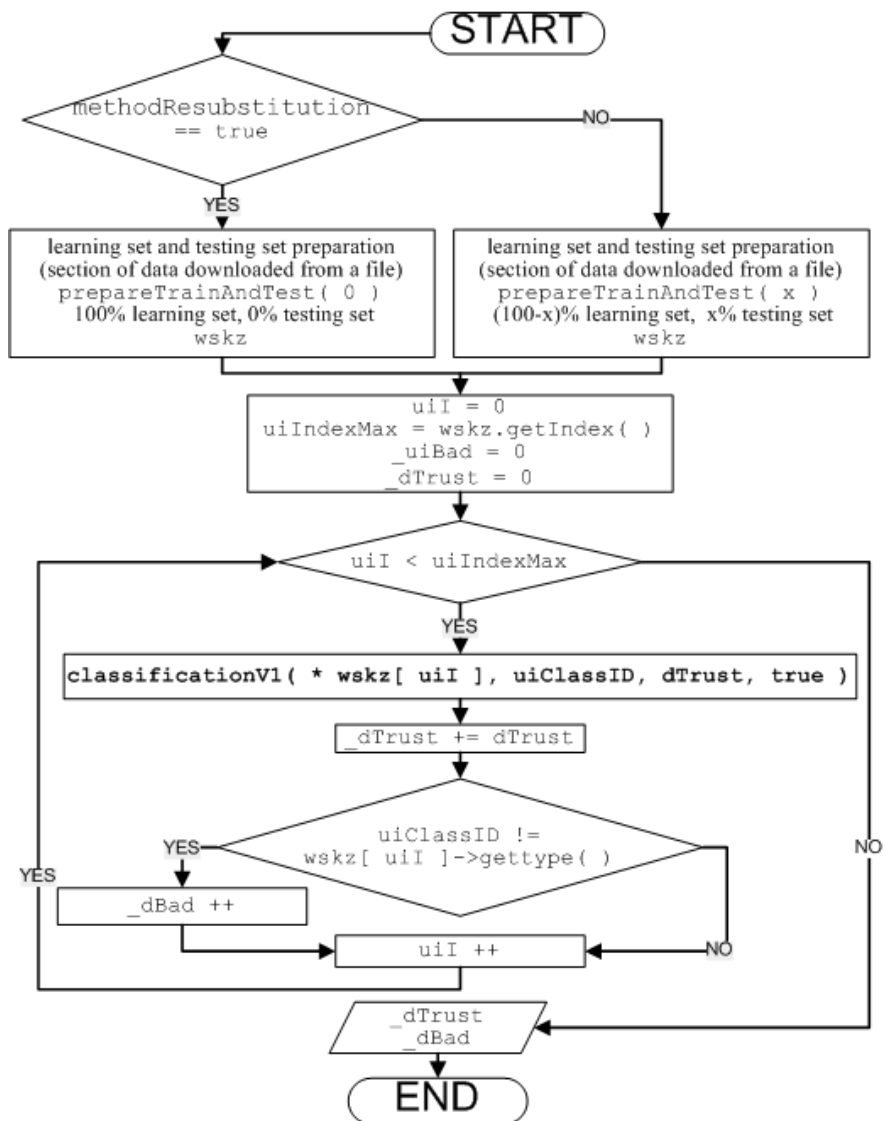


Fig. 3 The second stage of the algorithm's operation - determining the level of classification quality and the level of trust in the achieved results.

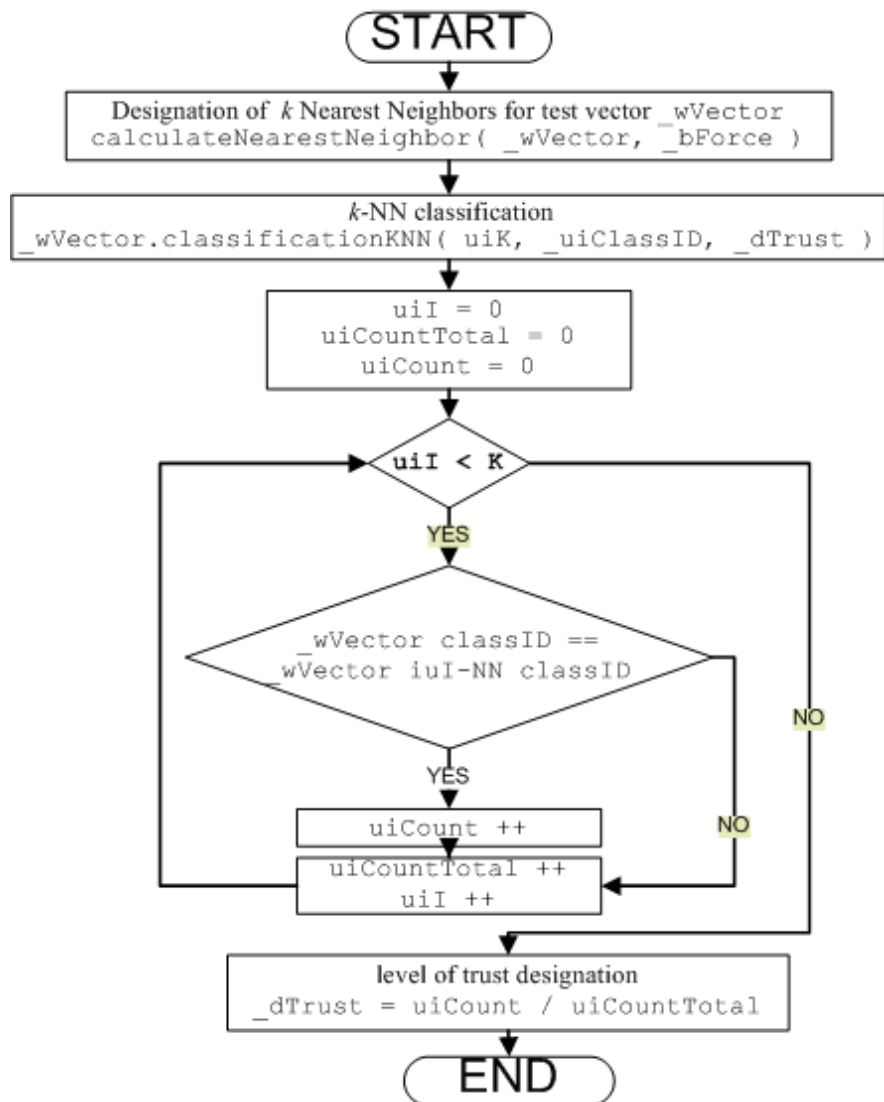


Fig. 4 The last stage of the algorithm's operation - classification with the k -NN method and the evaluation of the level of trust.

Table 1 Parameters of the sets used during the tests.

Name of the set	Number of classes	Number of features	Number of samples
Phoneme	2	5	5404
Satimage	6	36	6435
Waveform	3	21	5000

3 Verification of the Algorithm

The algorithm was implemented in C++ in Microsoft Visual Studio .NET 2008 environment. This allowed the authors to test the method in Windows environment with the use of a PC computer equipped with AMD TurionX2 Ultra DualCore processor 2,2GHz and 4GB of operating memory. The computation tests were conducted with the use of sets from the repository of the University of California in Irvine (Machine Learning Repository, University of California, Irvine) [11]. These tests are commonly used in literature. These are the following (Table. 1):

- *PHONEME* - data set created as a result of an analysis of separate syllables pronunciation (e.g. pa, ta, pan etc.); what was taken into account in this analysis was the type of a vowel pronunciation - nasal or oral;
- *SATIMAGE* - this data set was generated based on the analysis of satellite pictures supported with other methods of observation (radar data, topographic maps, data concerning agriculture). Classes determine a kind of soil or a type of cultivation;
- *WAVEFORM* - artificially generated data set, where each of the classes is created as a result of a combining 2 out of 3 sinusoids; for each attribute in a class noise is generated.

The coefficient k was determined only once for each testing set from the range of 1 - 21. The two following criteria were taken into account at the selection: the reliability of error calculation and the error itself; however, the first criterion was superior to the second one.

3.1 The PHONEME Testing Set

During the first stage of the algorithm's operation, the best coefficient k was determined for the *PHONEME* set. It was 1 ($k=1$); its determination took 211,703s. Then, tests evaluating the reliability for the determination of the 1-NN method's error were conducted with the best coefficient k (additionally, the time for error determination was measured). The mean values from 10 measurement series are presented below (Table. 2). In each measurement series, a new random division into the learning part and testing part was performed with proportions given in the table.

The tests indicated that the best coefficient k in terms of reliability was $k=1$. For all tests, it gave 100% reliability of error calculation. However, it should be

Table 2 The *PHONEME* set.

Division into the testing set and the learning set	Mean values		
	Error [%]	Reliability [%]	Time [s]
10 : 90	61,5197	100,0000	1,4454
25 : 75	42,9262	100,0000	3,5360
50 : 50	25,0407	100,0000	4,2624
75 : 25	15,5514	100,0000	3,6202
90 : 10	11,1852	100,0000	1,7386

emphasized that the great result of error calculation reliability for $k=1$ was connected with the increase in this error.

3.2 The *SATIMAGE* Testing Set

During the first stage, the algorithm determined the best coefficient k for the *SATIMAGE* set. It was 3 ($k=3$) and it took 499,428 s to determine it. With the best coefficient k , the tests evaluating reliability of the calculation of the 3-NN method's error were conducted (additionally, the time required for error determination was measured). Mean values from 10 measurement series are presented in the table (Table 3). During each measurement series, a new random division into the learning set and the testing set was performed with the proportions given in the table.

Table 3 The *SATIMAGE* set.

Division into the testing set and the learning set	Mean values		
	Error [%]	Reliability [%]	Time [s]
10 : 90	68,4856	84,3562	3,7090
25 : 75	35,8413	87,6937	8,4919
50 : 50	9,8042	92,9562	12,2850
75 : 25	5,6468	95,9121	9,1443
90 : 10	5,0389	96,4593	4,3405

Table 4 The *WAVEFORM* set.

Division into the testing set and the learning set	Mean values		
	Error [%]	Reliability [%]	Time [s]
10 : 90	73,2067	87,5870	1,7848
25 : 75	51,1333	90,6987	3,7824
50 : 50	35,5800	81,4181	5,7738
75 : 25	26,6080	79,4842	4,4978
90 : 10	22,7400	79,8829	2,1122

The results of the tests indicated that the best coefficient k in terms of reliability was $k=3$. It gave a high level of reliability of achieved results in all tests (even regardless from the aspect of the testing set overlearning) - min. 84%. It may also be observed that the best result was achieved with the overlearned set.

3.3 *The WAVEFORM Testing Set*

During the first stage of the algorithm's operation, the best coefficient k was determined for the *WAVEFORM* set. It was 21 ($k=21$); it took 247,498s to determine it. Then, tests were conducted to evaluate reliability of the calculation of 21-NN method's error with the best coefficient k (additionally, the time required for calculating the error was measured). Mean values from 10 measurement series are presented in the table below (Table. 4). In each measurement series, a new random division was performed into the learning set and the testing set with the proportions given in the table.

The tests indicated that the best coefficient in terms of reliability was $k=21$. It may be noticed that a large coefficient k adversely effects the evaluation of reliability of the achieved results. The level of reliability did not enable the authors to evaluate the error itself reliably

4 Conclusions

The conducted tests are only the beginning of works on the development of an algorithm steering between the speed of classification and classification error in relation to automatic evaluation of reliability of achieved results. It is clear that the selection of the parameters described above (in future works, the authors plan to determine their fixed values) may enable the user to reach a compromise between the global classification error and the number of objects which were classified with a determined level of reliability. In the next stage of the research, data sets describing network traffic (taken from literature or obtained from a functioning network by the authors) will be examined. The concept discussed in this study will constitute an idea of the engine for a network behavioural analysis system [4]. Further extensions of the application with the authors' solutions will make it possible to verify the usefulness of these solutions in computer networks.

References

- [1] Asuncion, A., Newman, D.J.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA (2007)
- [2] Stawska, Z., Józwiak, A., Sokołowska, B., Budzińska, K.: A multistage classifier based on distance measure and its use for detection of respiration pathology. In: Conference KOSYR 2001, Mińków k/Karpacza (2001)

- [3] Sierszeń, A., Sturgulewski, Ł.: Diffuse Network Behavior Analysis systems with the use of pattern recognition methods. In: Conference Information Systems Architecture and Technology, Wrocław 2010, pp. 223–234 (2010) ISBN: 978-83-7493-542-5
- [4] Sierszeń, A., Sturgulewski, Ł., Dubel, M., Marciniak, T., Wójciński, A.: Network Behavior-Analysis systems with the use of learning set and decision rules based on distance, Automatyka. AGH, Kraków (2010) ISSN: 1429-3447

Prototype Extraction of a Single-Class Area for the Condensed 1-NN Rule

Marcin Raniszewski

Abstract. The paper presents a new condensation algorithm based on the idea of a sample representativeness. For each sample in a dataset a representative measure is counted. Starting with samples with the highest value of the measure, each sample and all its voters (which constitute single-class area) are condensed in one averaged prototype-sample. The algorithm is tested on nine well-known datasets and compared with Jozwik's condensation methods.

1 Introduction

The main drawback of Nearest Neighbour method (NN) is a computational time and memory requirements for large training sets [4, 10]. The new sample is classified to the class of its nearest neighbour sample, hence there is a need to count the distances to all samples in a training set. To avoid counting distances to every sample from a training set, reduction and condensation methods have been created [11]. In this paper we refer to reduction algorithms as to the procedures which reduce the size of a training set simply by removing majority of samples. Only the samples most important (lying in the middle of single-class areas or near class boundaries) should be preserved in a reduced training set. On the other hand, condensation methods create new artificial samples (called prototypes) on the basis of existing samples and therefore they can be situated in points of a feature space, where there are no original samples.

In this article a new condensation method is presented. It is based on the idea of class representativeness. Samples with the high value of so called representative measure and the samples of their near neighbourhood, coming from the same class, are replaced by the prototype (average sample). In that way, the samples which constitute single-class region are replaced by one artificial sample representing the area.

Marcin Raniszewski

Technical University of Lodz, Computer Engineering Department

90-924 Lodz, ul. Stefanowskiego 18/22

e-mail: mrzanisz@kis.p.lodz.pl

In the second section of this paper well-known condensation method are mentioned (with the focus on Jozwik's procedure). Next, in the third section, the proposed condensation algorithm is described. At the end, test results are presented and some conclusions are formulated.

2 Well-Known Condensation Procedures

The historically first condensation algorithm is Chang's method [2]. The main idea is based on merging two closest prototypes into one. The result set is consistent with the original training set (1-NN rule correctly classifies all samples from the original training set operating with the result set): in every step of the procedure the consistency criterion has to be checked. It generates computational loads and long time of condensation for large datasets. Bezdek's slight improvement of Chang method, called Modified Chang Algorithm [1], does not eliminate these limitations. A further generalization of the idea is proposed in [7], where we replace the group of prototypes by one maintaining the consistency criterion.

All these methods generate consistent condensed datasets. But the consistency criterion does not work well on datasets with many outliers and atypical samples and does not guarantee to maintain a classification accuracy on sufficiently high level.

Jozwik in [3] does not use the consistency criterion. The dataset is simply divided into number of subsets (the number of subsets is defined by user) and each of the subset is replaced by the prototype representing dominant class in the subset (if class c is dominant in the subset, the prototype is gravity centre of all samples from the subset and it represents class c). To divide a set in Jozwik's method we have to find two farthest samples x and y from this set. The distance between these samples is called diameter of the set. Then samples which lie nearer to the sample x than to the sample y create the first subset and the rest of samples form the second subset. If we want to divide another subset, the subset with the greatest diameter is chosen. Subsets with samples from one class only (single-class subsets) are not further considered. There are two possible stop conditions of Jozwik's method: the first one occurs when the number of divisions is equal to that defined by user, the second is the case when there are no more subsets with more than one class.

Sanchez in [8] proposed three condensation methods based on Jozwik's algorithm, called RSP (Reduction by Space Partitioning). RSP1 and RSP2 slightly differ from Jozwik's method, while RSP3 is the same as Jozwik's procedure with the lack of defined by user number of divisions. Due to the fact that RSP3 results are the best, concerning classification accuracy, we will compare proposed in this paper algorithm with Jozwik's method applied with second stop condition.

Jozwik in [6] proposed a modified version of the method which consists in replacing the two furthest points of the set by two mutually furthest points. The algorithm of choosing points mutually the furthest is faster than searching for points with the greatest distance between them: we start with finding sample x_1 maximally distanced from the gravity centre of the set and then we find another sample x_2

maximally distanced from x_1 and so on until we find out that $x_{i-1} = x_{i+1}$ for some i . Then samples x_{i-1} and x_{i+1} form a pair of mutually furthest points. Pair of furthest points in the set is always a pair of mutually furthest points, but not each pair of mutually furthest points is the pair of furthest points in the set. The difference in choosing the furthest points strongly accelerates the Jozwik's condensation method but also causes slightly different condensed sets. Proposed in this paper algorithm is also compared with this variation of the Jozwik's method.

3 Proposed Method

The proposed in this paper method is based on representative measure which can be count for each sample from a training set. Before we define the mentioned measure let us define the idea of voter: a sample y is a voter of a sample x if y is from the same class as x and x is nearer to y than any other sample from different class. The representative measure of a sample is the number of its voters. It is easy to notice that a sample with the higher value of representative measure is more probable situated in the area of its class (Fig. 1). This sample and its voters create the area of their class. If we replace them with their gravity centre (prototype sample), it will represent the area well. This is the main concept used in proposed condensation procedure CRM (as a abbreviation of Condensation algorithm based on Representative Measure).

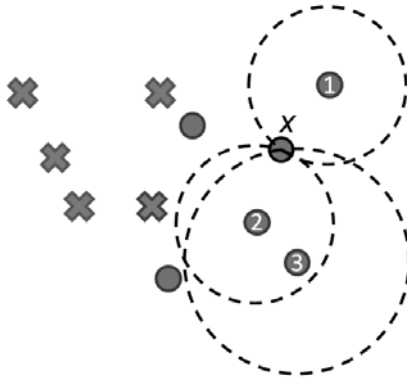


Fig. 1 The representative measure of the sample x is equal 3

Samples with the representative measure equal to 0 have no voters and they rather do not lie in the area of their class. In most cases they are outliers or atypical samples and they should not be taken to a condensed set.

The user can decide which areas are condensed by setting the value of minimal representative measure parameter ($rm_{min} \geq 1$). Only the samples with the representative measure higher or equal rm_{min} are considered by CRM.

The procedure CRM is very simple and consists of three steps:

1. Sort all samples in a training set according to decreasing values of their representative measure and set for all samples the flag "not condensed".
2. Take a sample from the training set (according to the arrangement done by the first step) with the flag "not condensed" and:
 - a. if its value of representative measure is higher or equal rm_{min} , condense it and its voters (count the gravity centre of the samples), set the flag "condensed" to the condensed samples and repeat step 2.
 - b. if its value of representative measure is less than rm_{min} go to step 3.
3. All created gravity centres form the condensed set.

The above algorithm requires few comments:

- selected sample and all its voters after condensation are no longer considered by the algorithm (the flag for them is set to "condensed" and only "not condensed" samples are further considered);
- only the first step of the algorithm is computationally burdensome but it is executed only once;
- after the construction of a condensed set for a fixed value of $rm_{min} \geq 2$ the algorithm can be simply continued (from step 2) for rm_{min} minus 1. In that way we can fast generate few condensed sets for decreasing values of rm_{min} ;
- result of the algorithm is repeatable.

4 Test Results

Eight real and one artificial datasets were used for testing: Liver Disorders (BUPA) [5], GLASS Identification [5], IRIS [5], PHONEME [9], PIMA Indians Diabetes [5], SAT [5], Wisconsin Diagnostic Breast Cancer (WDBC) (Diagnostic) [5], YEAST [5] and WAVEFORM (version1) [5] (Table 1).

Table 1 Datasets description

Dataset	No. of classes	No. of attributes	No. of samples
BUPA	2	6	345
GLASS	6	9	214
IRIS	3	4	150
PHONEME	2	5	5404
PIMA	2	8	768
SAT (training subset)	6	36	4435
SAT (testing subset)	6	36	2000
WAVEFORM	3	21	5000
WDBC	2	30	569
YEAST	10	8	1484

Table 2 Test results (classification accuracies in percentages)

Dataset	1-NN	Jozwik's1 (RSP3)	Jozwik's2	CRM1	CRM2	CRM3
BUPA	62,61	60,60	61,18	62,89	64,90	70,13
GLASS	71,56	68,36	69,76	67,84	65,94	64,30
IRIS	96,00	95,33	95,33	97,33	98,00	98,00
PHONEME	90,82	87,12	88,27	85,68	85,58	85,77
PIMA	67,20	63,42	62,76	67,98	73,45	73,18
SAT	89,45	89,50	89,15	89,15	89,40	89,55
WAVEFORM	77,40	77,82	78,26	81,96	82,52	82,52
WDBC	91,19	89,96	90,13	76,23	83,98	84,69
YEAST	53,04	51,08	52,10	54,72	57,72	56,59

Ten-fold cross-validation (with 1-NN) was used for each test (except for SAT dataset where we condensed whole training subset and tested it on testing subset).

The datasets were condensed by two algorithms: Jozwik's and proposed in this paper CRM. Two versions of Jozwik's method were tested (as mentioned in section 2):

- Jozwik's1 (RSP3) - division based on furthest points with no specified number of divisions (the algorithm ends when all subsets consist of single-class points);
- Jozwik's2 - division based on mutually furthest points with no specified number of divisions (the algorithm ends when all subsets consist of single-class points).

CRM were tested for three values of rm_{min} : 3, 2 and 1 (CRM3, CRM2 and CRM1 respectively). It is possible to obtain three condensed sets in a single CRM run (see section 3).

The results of classification accuracy are also compared with 1-NN operating on complete training set.

The best results for each dataset are highlighted in bold (Table 2 and Table 3).

Table 3 Test results (reduction rates in percentages)

Dataset	Jozwik's1 (RSP3)	Jozwik's2	CRM1	CRM2	CRM3
BUPA	36,55	35,59	64,83	78,13	85,73
GLASS	43,15	43,46	71,86	81,31	86,81
IRIS	84,37	84,37	88,67	88,74	88,96
PHONEME	70,97	70,66	81,32	84,31	86,90
PIMA	43,04	44,04	69,50	81,57	86,28
SAT	72,47	73,03	83,77	86,49	88,46
WAVEFORM	59,11	59,20	72,14	77,82	81,40
WDBC	80,08	79,98	89,85	91,58	92,74
YEAST	28,25	27,81	70,49	83,96	90,28

5 Discussion

As we can see in Table 2, we obtained better classification accuracies than that obtained on complete training set for few datasets (BUPA, IRIS, PIMA, SAT, WAVEFORM, YEAST).

Jozwik's1 (RSP3) and Jozwik's2 results are very similar, what is the consequence of only one slight difference in these procedures: whether the algorithm is based on the furthest or on the mutually furthest points. Jozwik's2 method is faster (selecting of mutually furthest points is less computationally complex than selecting furthest points) and provides the best classification accuracies on three datasets (GLASS, PHONEME, WDBC). The reduction rates of Jozwik's methods are rather poor (except for IRIS, SAT and WDBC datasets) - see Table 3.

Proposed in this paper CRM offers the best classification accuracies for six datasets (Table 2), especially for BUPA and YEAST datasets where the classification accuracy is much more higher than that obtained on complete datasets. Obviously we have 3 versions of CRM and as we can see only the CRM2 and CRM3 results are highly satisfying. CRM3 reduced all the datasets with the highest rate (from 81,40% to 92,74%). This reduction ratio is completely sufficient to optimize 1-NN. Reduction rates of CRM2 and CRM1 are lower due to the fact that for $a < b$ CRM a condensed set is always a superset of CRM b condensed set.

For all datasets the additional tests were made with CRM4 and CRM5 (not included in this paper) but for all tested datasets, the classification accuracies of these condensed sets were lower than that obtained by CRM3. Of course, the reduction rates were much higher than CRM3's.

A question is: what value of CRM's rm_{min} should we used in real applications? The recommended values are 2 and 3. User should perform tests and choose the better alternative. It is possible that for some databases CRM1 will result in best classification accuracies, so if CRM2 or CRM3 results are not satisfactory, we should check CRM1 especially taking into account that all condensed sets CRM3, CRM2 and CRM1 can be generated in a single run of CRM.

Running time of all tested algorithms was very short (only few seconds for datasets with thousands of samples like PHONEME, SAT and WAVEFORM).

6 Conclusions

The proposed condensation method is based on replacing single-class areas by one prototype. It uses a representative measure: sample in the middle of single-class area has high value of representative measure and is condensed with all its voters (other samples from the area) to one gravity centre. The procedure ends if there is no sample with the representative measure higher or equal to a predefined minimal representative measure.

The method has the following features:

- satisfactory high classification accuracy (for some datasets even better than that obtained on complete dataset);

- high reduction rate;
- results are repeatable;
- one parameter: minimal representative measure (rm_{min});
- the possibility of creating ascending series of condensed sets in a single run (it is recommended to generate two condensed sets for rm_{min} equal to 3 and 2 and choose the better set);
- very short time of condensation phase.

References

- [1] Bezdek, J.C., Reichherzer, T.R., Lim, G.S., Attikiouzel, Y.: Multiple-prototype classifier design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 28(1), 67–79 (1998)
- [2] Chang, C.L.: Finding prototypes for nearest neighbor classifiers. *IEEE Transactions on Computers* C-23(11), 1179–1184 (1974)
- [3] Chen, C.H., Jozwik, A.: A sample set condensation algorithm for the class sensitive artificial neural network. *Pattern Recognition Letters* 17(8), 819–823 (1996)
- [4] Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*, 2nd edn. John Wiley & Sons, Inc., Chichester (2001)
- [5] Frank, A., Asuncion, A.: *UCI Machine Learning Repository*. University of California, School of Information and Computer Science, Irvine, CA (2010), <http://archive.ics.uci.edu/ml>
- [6] Jozwik, A., Kies, P.: Reference set size reduction for 1-NN rule based on finding mutually nearest and mutually furthest pairs of points. *Computer Recognition Systems, Advances in Soft Computing* 30, 195–202 (2005)
- [7] Mollineda, R.A., Ferri, F.J., Vidal, E.: An efficient prototype merging strategy for the condensed 1-NN rule through class-conditional hierarchical clustering. *Pattern Recognition* 35(12), 2771–2782 (2002)
- [8] Sanchez, J.S.: High training set size reduction by space partitioning and prototype abstraction. *Pattern Recognition* 37(7), 1561–1564 (2004)
- [9] The ELENA Project Real Databases, <http://www.dice.ucl.ac.be/neural-nets/Research/Projects/ELENA/databases/REAL/>
- [10] Theodoridis, S., Koutroumbas, K.: *Pattern Recognition*, 3rd edn. Academic Press - Elsevier, USA (2006)
- [11] Wilson, D.R., Martinez, T.R.: Reduction techniques for instance-based learning algorithms. *Machine Learning* 38(3), 257–286 (2000)

Robust Nonparametric Regression with Output in $SO(3)$

Grzegorz Jabłoński*

Abstract. In this paper the implementation of robust nonparametric kernel estimator for $SO(3)$ group is presented. We propose the conjugate gradient method for solving the optimization problem, which arises during computation of the estimator. Finally an experiment with database of toy figure images together with their rotations is conducted.

1 Introduction

The need for structured output learning comes from more sophisticated problems in the area of machine learning. Basic models offer only a limited possibility to use a priori knowledge about the structure of data. Standard classification and regression are not sufficient for some real life applications. Structured output learning tries to use the structure of input and output data. Because of that, it is successfully employed in areas like computational biology, natural language processing or graph theory [7]. In [4] it is stated that one can see regression with output space restricted to Riemannian manifold as structured output learning. The idea behind that is to think of Riemannian manifolds as isometrically embedded in the Euclidean space.

In this paper firstly we present results of [4]. Next we use [5] to implement Hein's kernel estimator on $SO(3)$ manifold. The main focus on $SO(3)$ manifold includes definitions and the use of distance and geodesics on $SO(3)$. We propose the conjugate gradient method based on [2] (instead of steepest descent) for solving optimization problem. This step is crucial in any structure output estimation algorithm. We show that for an experimental data the conjugate gradient method has a faster convergence rate. Finally we use database from [6] to test the algorithm. In the experiment a rotation of a toy figure is estimated basing on some previous images.

Grzegorz Jabłoński

Institute of Computer Science, Jagiellonian University, ul. Łojasiewicza 6 Krakow

e-mail: grzegorz.jablonski@uj.edu.pl

* I would like to thank dr Matthias Hein and dr Igor Podolak for their helpful comments and ideas.

2 Background

The goal is to learn optimal mapping $\phi : M \rightarrow N$, where M and N are metric spaces. d_M and d_N denote metric of M and N respectively. Both M and N are separable and given data (X_i, Y_i) are i.i.d. samples from probability measure P on $M \times N$. Definitions and formulas in this section comes from [4].

Like in Euclidean space we will use loss functions penalizing the distance between predicted output and desired output. Loss function

$$L(y, \phi(x)) = \Gamma(d_N(y, \phi(x))), y \in N, x \in M \quad (1)$$

where $\Gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a function from a restricted family. Γ has to be (α, s) -bounded (we omit the exact definition from [4] due to paper length restrictions), if one wants to show consistency of a considered estimator and existence of optimal solution [4]. Hein defined the family of kernel estimators with metric-space valued output. Let $(X_i, Y_i)_{i=1}^l$ be the sample with $X_i \in M$ and $Y_i \in N$. The metric-space-valued kernel estimator $\phi_l : M \rightarrow N$ from metric space M to metric space N is defined for all $x \in M$ as

$$\phi_l(x) = \operatorname{argmin}_{q \in N} \frac{1}{l} \sum_{i=1}^l \Gamma(d_N(q, Y_i)) k_h(d_M(x, X_i)) \quad (2)$$

where $\Gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is (α, s) bounded, $k : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, $k_h = \frac{1}{h} k(x/h)$ and k is a kernel function.

One can rewrite the formula [2] into functional $F(q)$ for $q \in N$, which could be optimized. Its solution will give us an optimal element for the kernel estimator $\phi_l(x)$

$$F(q) = \sum_{i=1}^l w_i \Gamma(d_N(p, Y_i)) \quad (3)$$

where $w_i = k_h(d_M(x, X_i))$. Covariant gradient of the functional F at a point q is given by:

$$\nabla F|_q = \sum_{i=1}^l w_i \Gamma'(d_N(p, Y_i)) v_i \quad (4)$$

where $v_i \in T_q N$ is a tangent vector at q with $\|v_i\|_{T_q N} = 1$ given by the tangent vector at q of the minimizing geodesic from Y_i to q (pointing away from Y_i) [4].

The algorithm in [4] used to find minimum is based on the gradient-descent optimization algorithm with Armijo rule [1]. It starts from random point on the manifold and iterates until specific criteria are fulfilled:

```

 $q_0 :=$  random point on manifold
while(stopping criteria not fulfilled)do
  compute gradient  $F|_q$  at point  $q_k$ 
  compute step size  $\alpha$  based on Armijo rule
   $q_{k+1} := \exp(\alpha \nabla F|_q)$ 

```

In our implementation condition $\|q_{k+1} - q_k\| < \varepsilon$ was used as a stopping criteria. However, one can also use the norm of the gradient.

3 $SO(3)$ Group

$SO(3)$ group is identified with a group of all rotations about the origin of three-dimensional Euclidean space \mathbb{R}^3 under the operation of composition. Rotation about the origin is a linear transformation which preserves distance of points and orientation of space. For detailed description of $SO(3)$ we refer to [5]. Most of the definitions and equations used in this section were taken from Moakher work. Let $GL(3)$ be group of 3-by-3 real, non-singular matrices; the Lie group of rotations is:

$$SO(3) = \{R \in GL(3) | R^T R = I \wedge \det R = 1\}, \quad (5)$$

where I is identity transformation and the superscript T denotes matrix transpose. Corresponding Lie algebra $so(3)$ is the space of skew-symmetric matrices

$$so(3) = \{A \in gl(3) | A^T = -A\} \quad (6)$$

where $gl(3)$ denotes the space of linear transformations in \mathbb{R}^3 .

3.1 *Logarithm and Exponential of Matrix*

Theoretical definitions of exponential and logarithm of matrix are useless in terms of computer applications. Thus we will use following Rodrigues' formulas for 3-by-3 matrices

$$\exp A = \begin{cases} I & \text{if } a = 0 \\ I + \frac{\sin a}{a} A + \frac{1 - \cos a}{a^2} A^2 & \text{if } a \neq 0 \end{cases} \quad (7)$$

where A is a skew-symmetric matrix, I denotes identity matrix and $a = \sqrt{\frac{1}{2} \text{tr}(A^T A)}$. The principal logarithm for a matrix $A \in SO(3)$ is given by

$$\log A = \begin{cases} 0 & \text{if } \omega = 0 \\ \frac{\omega}{2 \sin \omega} (A - A^T) & \text{if } \omega \neq 0 \end{cases} \quad (8)$$

where ω satisfies $\text{tr} A = 1 + 2 \cos \omega$ and $|\omega| < \pi$.

3.2 *Logarithm and Exponential Mappings*

We present a formula which describes tangent space at a point $p \in SO(3)$:

$$T_p SO(3) \cong \{pk | k \in \mathbb{K}(3, \mathbb{R})\} \quad (9)$$

where $\mathbb{K}(3, \mathbb{R})$ denotes the set of $n \times n$ skew-symmetric matrices.

From above one can easily conclude that tangent space at identity rotation is equal to set of all skew-symmetric matrices. For our purposes we will first "move" our point p to identity rotation, then compute logarithm and finally "go back" with computed logarithm to the initial point. It is possible due to the isomorphism between tangent spaces at various points.

Simplified exponential and logarithm mappings in $SO(3)$ are given by:

$$\text{Exp}_p(q) = p \exp(q) \quad (10)$$

where \log and \exp denotes matrix logarithm and exponential, $q \in so(3)$, $p \in SO(3)$.

$$\text{Log}_p(q) = \log(p^T q) \quad (11)$$

where $\log(p^T q) \in so(3)$ Instead of moving the projection of rotation to a tangent space at point p we will just use its projection onto $so(3)$ (Lie algebra). Following that pattern, all operations are performed in the tangent space at identity rotation ($so(3)$). Only the final value is transformed to a vector at $T_p M$. Using this equations we perform less transformations and gain better accuracy than working in tangent spaces of all rotations.

Above definitions provide straightforward formulas for geodesics on $SO(3)$ manifold. Direction in which we are moving is a linear space, so we could use multi-operation in this space. Geodesic emanating from p in direction q is given by $G(t) = p \exp(tq)$, where $t \in [0, 1]$, $q \in so(3)$.

3.3 Riemannian Distance

The distance functions will be defined in $SO(3)$. Metric used in our implementation comes from [5]. We use the Riemannian structure on $SO(3)$. Riemannian distance is:

$$d_R(R_1, R_2) = \frac{1}{\sqrt{2}} \|\text{Log}(R_1^T R_2)\| \quad (12)$$

where $\|\cdot\|_F$ is the Frobenius norm induced by the Euclidean inner product, defined by $\langle R_1, R_2 \rangle = \text{tr}(R_1^T R_2)$. d_R is the length of the geodesic (shortest curve lying in $SO(3)$) connecting R_1 and R_2 . In the conjugate gradient method also non-Euclidean product is applied: $\langle R_1, R_2 \rangle = \frac{1}{2} \text{tr}(R_1^T R_2)$. Metric induced by such product agree with Frobenius norm up to a constant factor $\frac{1}{2}$.

4 Implementation

Since we can compute logarithm and matrix mappings, the algorithm for finding kernel estimator may be presented. First we insert eq. [1] into covariant gradient formula in eq. [4]. Then replace tangent vectors at given point q with projections onto tangent space at q . Covariant gradient on $SO(3)$ is

$$\nabla F|_q = \sum_{i=1}^l w_i \Gamma'(d_R(q, Y_i)) v_i = \sum_{i=1}^l w_i \Gamma' \left(\frac{1}{\sqrt{2}} \| \text{Log}(q^T Y_i) \| \right) \text{Log}(q^T Y_i) \quad (13)$$

We want to find a minimum of the nonlinear function f . An outline for the nonlinear conjugate gradient method goes as follows: at the k th iteration minimize function f along the H_{k-1} direction and compute new direction H_k which is a combination of the gradient in x_k and the old search direction. In an idealized version of nonlinear conjugate gradient the old and new search direction must be conjugate with respect to the Hessian of f . Here is the formula for the new search direction

$$H_{k+1} = -G_{k+1} + \gamma_k \tau H_k, \text{ where } \gamma_k = \frac{-\text{Hess}f(G_k, \tau H_{k-1})}{\text{Hess}f(\tau H_{k-1}, \tau H_{k-1})} \quad (14)$$

and τ denotes parallel transportation of H_{k-1} to x_k . We have simple formula for τH_k , which is later applied in the code in this section. It is possible to approximate γ_k without computation of Hessian and with no loss of functionality. The idea behind this simplification is to capture the second derivative information by computing only first derivatives [2]. One obtains formulas:

$$\text{Polak-Ribiere: } \gamma_k = \frac{\langle G_k - \tau G_{k-1}, G_k \rangle}{\langle G_{k-1}, G_{k-1} \rangle} \quad (15)$$

$$\text{Fletcher-Reeves: } \gamma_k = \frac{\langle G_k, G_k \rangle}{\langle G_{k-1}, G_{k-1} \rangle} \quad (16)$$

Unfortunately, we do not have formula for τG_k , so in practice we use $\tau G_k := 0$ or $\tau G_k := G_k$. It does not affect the superlinear convergence of the method [2].

We use the fact that $SO(3)$ is a Stiefel manifold of 3-by-3 matrices and therefore the conjugate gradient on the Stiefel manifold could be used. An algorithm from [2] is implemented. We do some changes:

- computation of G_k (gradient of F in iteration k , in our implementation $\nabla F|_q$) is computed as in equation [3]
- minimization of $F(Y_k(t))$ over t is performed using the golden section search [3] (in our implementation $F(q_k(t))$).

The conjugate gradient method on $SO(3)$ is:

```

input:  $(X_i, Y_i)$ ,  $x$ , where  $X_i$  - input space,  $Y_i$  - rotation,  $x$  point to estimate
 $w_i := k_h(d_M(x, X_i))$ ,  $l :=$  number of rotations in training set
 $q_0 :=$  random matrix from  $SO(3)$ 
 $FunVal_0 := \sum_{i=1}^l w_i \Gamma(\text{Log}(q_0^T Y_i))$ 
 $\nabla F|_{q_0} := \sum_{i=1}^l w_i \Gamma' \left( \frac{1}{\sqrt{2}} \| \text{Log}(q_0^T Y_i) \| \right) \text{Log}(q_0^T Y_i)$ 
for  $k := 0, 1, \dots$  do
   $QR :=$  compact QR decomposition of  $(I - Y_k Y_k^T) H_k$ 
   $A := q_k^T H_k$ 
  minimize  $F(q_k(t))$  over  $t$  using the golden section search
  computing  $q_k(t)$  as follows

```

$$\begin{aligned} \begin{pmatrix} M(t) \\ N(t) \end{pmatrix} &:= \exp t \begin{pmatrix} A & -R^T \\ R & 0 \end{pmatrix} \begin{pmatrix} I_p \\ 0 \end{pmatrix} \\ q_k(t) &:= q_k M(t) + QN(t) \\ t_k &:= t_{min} \\ q_{k+1} &:= q_k(t_k) \text{ minimizing argument} \\ \nabla F|_{q_{k+1}} &:= \sum_{i=1}^l w_i \Gamma' \left(\frac{1}{\sqrt{2}} \| \text{Log}(q_{k+1}^T Y_i) \| \right) \text{Log}(q_{k+1}^T Y_i) \\ \tau H_k &:= H_k M(t_k) - q_k R^T N(t_k) \\ \gamma_k &:= \frac{\langle \nabla F|_{k+1}, \nabla F|_{k+1} \rangle}{\langle \nabla F|_k, \nabla F|_k \rangle} \\ H_{k+1} &:= -\nabla F|_{k+1} + \gamma_k \tau H_k \\ \text{if } k+1 \bmod 3 = 0 \text{ then} & \\ \quad H_{k+1} &:= -\nabla F|_{k+1} \\ \text{if } \|q_k - q_{k+1}\| < \varepsilon \text{ then} & \\ \quad \text{exit for loop and return } q_k \text{ as a result} & \end{aligned}$$

5 Experiments

We want to rate the conjugate gradient method. We will demonstrate that a conjugate gradient method, proposed in this article, is faster than the steepest descent method. Interesting part was to generate curves in $SO(3)$. For this we used a special representation of rotation. Every rotation could be seen as an axis of rotation and his angle. The axis of rotation could be represented as a point on sphere (the actual axis will be given by the line connecting origin with point on sphere), and angle will be the distance from origin. Because angle could be in $[0, 2\pi)$ range thus rotations will be represented as a curve inside the sphere with radius 2π . As ground-truth for the experiment, we choose a curve given in spherical coordinates as $r(t) = 1 + \sin(t), \phi(t) = 10t^2, \theta(t) = 0.5\pi t + 1.2\pi \sin(\pi t)$. t was sampled uniformly from $[0; 1]$ range. After curve was generated, noise with normal distribution was added. Then every point was transformed to the closest proper rotation and t was given as input space (denoted as M in sec. 2) with corresponding rotation. Standard deviation of noise level used in experiment equals 0.01 or 0.05. The kernel used had the form: $k(\frac{x}{h}) = 1 - \frac{x}{h}$. Parameter h for kernel was chosen using 5-cross validation. Evaluation was based on rotations independent of input data. We have also averaged every set of parameters by running all experiments 3 times. To rate convergence we were estimating 100 rotations based on data with additive noise (standard deviation 0.05). In our experiment conjugate gradient method reach desired tolerance after at most 6 iterations, whereas steepest descent method needs at least 2 more iterations. We juxtapose the average number of iterations executed by both methods during estimating 100 rotations with various noise level - see fig. 2. There is two types of input data prepared. First set consist of rotations without outliers, whereas second has 20% of outliers (noise added to outliers had standard deviation equal to 0.7). For the data without outliers $L2$ cost function is applied (due to its robustness) while $L1$ is used for the second set. We have also compared the total amount of time needed to compute these experiments. One can see that the conjugate gradient reach its minimum at least two times

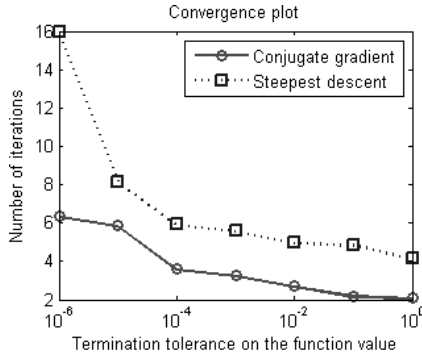


Fig. 1 Convergence of both methods. Values of $\|q_k - q_{k+1}\|$ (change of function value between iterations) when algorithm is stopped are on x-axis. Number of iterations needed to reach this value is presented on y-axis.

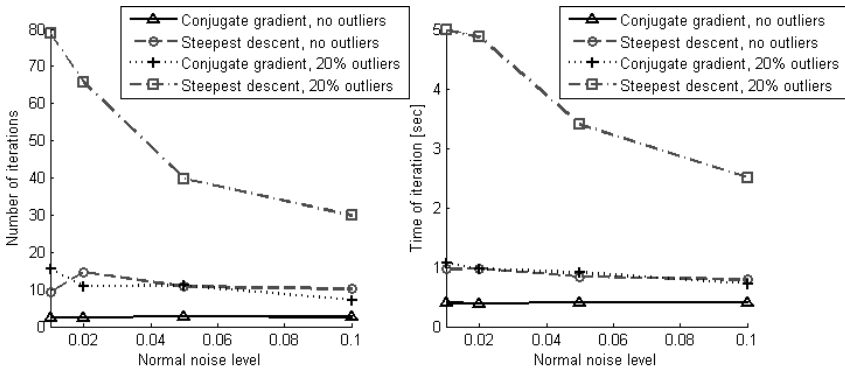


Fig. 2 Comparison of iteration number and running time of the two presented methods. Standard noise deviation is on x-axis. Average number of iterations (average running time of one iteration on right) necessary to find optimum value is on the y-axis.

faster than the steepest descent method. It is of great importance to notice that the conjugate gradient method does not need an extra time for additional computations. We could benefit from lower number of iterations without any visible loss caused by computation of additional parameters for the conjugate gradient method.

In tab. 1 we present the results of conducted experiments. As expected, L2 and Hueber loss presented better robustness, while L1 works better in cases with more outliers. Loss function with the highest accuracy within a group of parameters is highlighted. It is also important that the average error of conjugate method is very close to the error of steepest descent method.

Table 1 Results of experiment. The best values in every row are highlighted.

Error for steepest descent method (noise level 0.01)				
	Loss	L1	L2	Huber loss
100	0% outliers	0.0223 ± 0.0019	0.0157 ± 0.0020	0.0209 ± 0.0020
rotations	20% outliers	0.0343 ± 0.0089	0.0401 ± 0.0064	0.0346 ± 0.0063
500	0% outliers	0.0079 ± 0.0018	0.0058 ± 0.0007	0.0068 ± 0.0029
rotations	20% outliers	0.0077 ± 0.0011	0.0087 ± 0.0010	0.0097 ± 0.0031
Error for conjugate gradient method (noise level 0.01)				
	Loss	L1	L2	Huber loss
100	0% outliers	0.0165 ± 0.0021	0.0115 ± 0.0018	0.0178 ± 0.0024
rotations	20% outliers	0.0247 ± 0.0023	0.0485 ± 0.0110	0.0287 ± 0.0016
500	0% outliers	0.0069 ± 0.0010	0.0037 ± 0.0004	0.0056 ± 0.0001
rotations	20% outliers	0.0053 ± 0.0011	0.0124 ± 0.0015	0.0064 ± 0.0005
Error for steepest descent method (noise level 0.05)				
	Loss	L1	L2	Huber loss
100	0% outliers	0.0236 ± 0.0054	0.0175 ± 0.0033	0.0237 ± 0.0034
rotations	20% outliers	0.0401 ± 0.0103	0.1071 ± 0.0708	0.0457 ± 0.0132
500	0% outliers	0.0069 ± 0.0013	0.0047 ± 0.0004	0.0079 ± 0.0012
rotations	20% outliers	0.0072 ± 0.0009	0.0100 ± 0.0019	0.0093 ± 0.0010
Error for conjugate gradient method (noise level 0.05)				
	Loss	L1	L2	Huber loss
100	0% outliers	0.0198 ± 0.0019	0.0117 ± 0.0015	0.0190 ± 0.0056
rotations	20% outliers	0.0227 ± 0.0075	0.0416 ± 0.0028	0.0243 ± 0.0053
500	0% outliers	0.0056 ± 0.0009	0.0041 ± 0.0005	0.0057 ± 0.0015
rotations	20% outliers	0.0065 ± 0.0009	0.0100 ± 0.0004	0.0065 ± 0.0014

**Fig. 3** Original image of figure (left) and with applied edge filter (right).

5.1 Toy Figure Experiment

We have employed a database of 2500 grayscale images from [6] of toy figure to study our method. The toy figure, Tom, is photographed from upper half of the viewing sphere. Images are taken every 3.6 degree in horizontal and vertical direction. We would like to estimate the angle of an object on image using previously known images with angles. To apply the conjugate gradient method one needs metric in the input space. In this case it is a space of images sized 128x128. To simplify the computation first we apply canny filter on all images to detect edges. Such images are treated as sets of points. A modified Hausdorff distance between sets is used to compute the distance between two images (it may not be metric, but in practice gives better results). Training data consisted of 100 randomly chosen images (4% of all images). Images presented in the article were also chosen randomly from all images not used in training set. We have estimated rotations of all images in database. MSE equals 0.5277, mean error 0.3511. However one can see in the histogram that most values of error are in (0;0.3) range. In our opinion such mistakes were hardly to distinguish in plots of images with rotations.

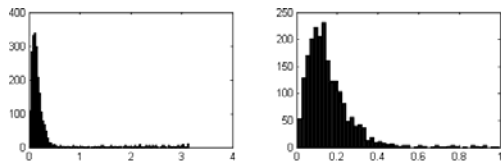


Fig. 4 Distribution of error. We have estimated rotations of all images in database using 100 (4%) randomly chosen images. Distribution of error is presented in the first plot, in the second plot distribution of error smaller than 1 (90% of all data).

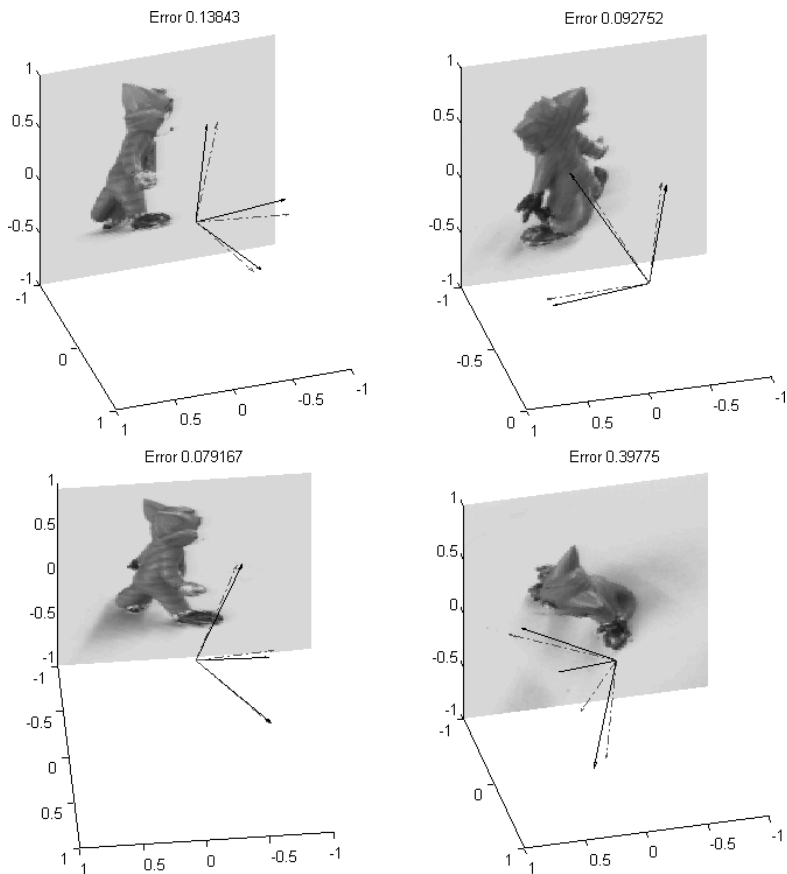


Fig. 5 Images of figure with estimated (solid line) and original (dashed line) rotations. Distance between these rotations is displayed as error at the top of the plots.

6 Summary

We demonstrated practical application of Hein's kernel estimator. Also implementation of conjugate gradient method based on [2] to solve optimization problem from [4] was presented along with results of comparison with the steepest descent method. The conjugate gradient implementation proposed in this article is faster and more suitable for real life applications. The code of this framework is available freely from www.ii.uj.edu.pl/~jablonsk/so3/. In the future we would like to extend our study to more general cases like orthogonal group or Stiefel manifold.

References

- [1] Bertsekas, D.: *Nonlinear Programming*. Athena Scientific, Belmont (1999)
- [2] Edelman, A., Arias, T., Smith, S.: The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications* 20 (2) (1999)
- [3] Forsythe, G., Malcolm, M., Moler, C.: *Computer Methods for Mathematical Computations*. Prentice-Hall, Englewood Cliffs (1977)
- [4] Hein, M.: Robust Nonparametric Regression with Metric-Space valued Output. *Advances in Neural Information Processing Systems* 22 (2009)
- [5] Moakher, M.: Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications* 24 (1) (2002)
- [6] Peters, G., Zitova, B., Malsburg, C.: How to Measure the Pose Robustness of Object Views. *Image and Vision Computing* 20 (4) (2002)
- [7] Ricci, E., De Bie, T., Cristianini, N.: Magic Moments for Structured Output Prediction. *Journal of Machine Learning Research* 9 (2008)

Time Series Prediction with Periodic Kernels

Marcin Michalak

Abstract. This short article presents the new algorithm of time series prediction: *PerKE*. It implements the kernel regression for the time series directly without any data transformation. This method is based on the new type of kernel function — periodic kernel function — which two examples are also introduced in this paper. This new algorithm belongs to the group of semiparametric methods as it needs the initial step that separate the trend from the original time series.

Keywords: time series prediction, regression, kernel methods, periodic kernel functions, semiparametric methods.

1 Introduction

Estimation of the regression function is the one of the fundamental task of machine learning. Its essence is to describe dependencies in the analyzed data set. These dependencies may be overt (for the expert) or hidden. In the first case evaluation of the regression function is often divided into two steps: the first step consists in the choice of the class of the regression function, which parameters are evaluated in the second step. This kind of regression is called parametric. Nonparametric estimators do not make any assumptions about the functional form of dependencies between independent and dependent variables. In other words, nonparametric estimator approximates values but does not try to explain the nature of the dependence.

Time series represent a wide group of present problems like financial, geological or transportation. Simultaneously, science offers a broad range of time series prediction methods. In spite of the individual character of time series, their prediction may be performed with the usage of regression methods, like kernel estimators.

Marcin Michalak

Central Mining Institute, Plac Gwarkow 1, 40-166 Katowice, Poland

e-mail: Marcin.Michalak@gig.eu

and

Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland

e-mail: Marcin.Michalak@polsl.pl

In the previous works the author presented kernel approach to the time series prediction [8, 9] but these method needed the transformation of the original time series into the new space.

This short paper introduces the *PerKE* (Periodic Kernel Estimator) the new algorithm for time series prediction. It belongs to the wide group of kernel estimators but is dedicated for time series. The essential part of the *PerKE* is the definition of the periodic kernel. Two function that satisfy condition of this type of kernel are also presented.

This paper is organized as follows: next section describes several methods of time series prediction (decomposition and *SARIMA*) and some foundation of non-parametrical kernel regression. Then the general definition of periodic kernel is formed and two periodic kernels are introduced. Afterwards the *PerKE* algorithm is described. Next section contains the description of analyzed data sets and results of experiments. The paper ends with some conclusions and perspectives of further works.

2 Prediction and Regression Models

2.1 *ARIMA (SARIMA) Models*

The *SARIMA* model is the generalization of the Box and Jenkins *ARIMA* model (*AutoRegressive Integrated Moving Average*) [2]. This model is the connection of three more simply models: autoregression (*AR*), moving average (*MA*) and integration (*I*). For the time series x the lag operator B is defined in the following way: $Bx_t = x_{t-1}$. Then the autoregressive model of the order P may be defined as: $x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + a_t$ where a_t is the white noise. This equation may be also presented with the usage of lag operator: $(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)x_t = a_t$.

In the *MA* models the value of the time series depends on random component a_t and its q delays as follows: $x_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$ or as $(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)a_t = x_t$ The final model *ARIMA* is used when the time series is not stationary. Then the d operation of its differentiation is performed that is described by the component $(1 - B)^d$ in the final equation. The full *ARIMA*(p, d, q) model takes the form:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)(1 - B)^d x_t = a_t (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$$

The *SARIMA* model (*Seasonal ARIMA*) is dedicated for time series that have strong periodic fluctuations. If s is the seasonal delay the model is described as *SARIMA*(p, d, q)(P, D, Q) ^{s} where P is the order of seasonal autoregression $(1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{Ps})$, Q is the order of seasonal moving average $(1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_p B^{Ps})$ and D is the order of seasonal integration $(1 - B^s - B^{2s} - \dots - B^{Ps})$.

2.2 Decomposition Method

The decomposition method tries to separate as much as possible components of the analyzed time series. As the most important components the following should be mentioned:

- trend component (T) that reflects the long time characteristic of the time series,
- seasonal component (S) that reflects the periodic changes of the time series,
- cyclical component (C) that describes repeated but non-periodic changes of the time series,
- irregular (random) component (e).

There are two models of decompositions that differ the way of components aggregation. First is called additive and the second is called multiplicative.

2.3 Kernel Estimators

Kernel estimators are ones of the simplest and probably the clearest examples of nonparametric estimators. For example, the Nadaraya–Watson estimator is defined in the following way [10, 17]:

$$\tilde{f}(x) = \frac{\sum_{i=1}^n y_i K((x - x_i)/h)}{\sum_{i=1}^n K((x - x_i)/h)} \quad (1)$$

where $\tilde{f}(x)$ means the estimator of the $f(x)$ value, n is a number of train pairs (x, y) , K is a kernel function and h is the smoothing parameter. This estimator assumes that independent (X) and dependent (Y) variables are random variables and the value of the regression function is an approximation of the conditional expected value of the dependent variable Y upon the condition, that the independent variable X took the value x :

$$\tilde{f}(x) = E(Y|X = x) \quad (2)$$

But more simply interpretation may sound that the estimator of the value $f(x)$ is the weighted average of observed values y_i . Similar kernel estimators are: Gasser–Muller [6], Priestley – Chao [16], Stone–Fan [4].

The function must meet some criterions [11] to be used as the kernel function:

1. $\int_{\mathbb{R}} K(u) du = 1$
2. $\forall x \in \mathbb{R} \quad K(x) = K(-x)$
3. $\int_{\mathbb{R}} uK(u) du = 0$
4. $\forall x \in \mathbb{R} \quad K(0) \geq K(x)$
5. $\int_{\mathbb{R}} u^2 K(u) du < \infty$

One of the most popular is the Epanechnikov kernel [3]:

$$K(x) = \frac{3}{4} (1 - x^2) I(-1 < x < 1) \quad (3)$$

Table 1 Popular kernel functions.

Uniform	$K(x) = \frac{1}{2}I(-1 < x < 1)$
Triangular	$K(x) = (1 - x)I(-1 < x < 1)$
Biweight	$K(x) = \frac{15}{16}(1 - u^2)I(-1 < x < 1)$
Gaussian	$K(x) = \frac{1}{\sqrt{2\pi}} \exp -u^2/2$

where $I(A)$ means the indicator of the set A . Other popular kernel function are presented in the Table 1.

The second step of the kernel estimator creating is the selection of the smoothing parameter h . As it is described in [12] and [15], the selection of h is more important than the selection of the kernel function. Small values of the h cause that the estimator fits data too much. Big values of this parameter h lead to the estimator that oversmooths dependencies in the analysed set.

The most popular method of evaluation the h parametr is the analysis of the approximation of the Mean Integrated Square Root Error (*MISE*), that leads to the following expression:

$$h_0 = 1.06 \min(\tilde{\sigma}, \tilde{R}/1.34)n^{-\frac{1}{5}} \quad (4)$$

where $\tilde{\sigma}$ is the standard deviation from the train set and \tilde{R} means interquartile range from the train set. Details of derivations can be found in [12]. More advanced methods of h evaluation can be found in [4, 5, 13, 14, 15].

2.4 Semiparametric Regression

One of the nonparametrical regression modification is the model of two step regression [1, 7]. First step is called parametrical and the second is called nonparametrical. This model is also called semiparametrical. Generarly, in the first step we assume some kind of parametrical form of the regression function. It may be linear, polynomial, exponential, but described with the finite set of parameters Θ . Then the estimation of unknown parameters Θ is performed with one of popular methods (least squares, *EM*, etc...). Afterwards the vector of regression function parameters estimators $\tilde{\Theta}$ is obtained that is used for the parametrical regression of the unknown function: $\tilde{f}(x, \tilde{\Theta})$. This step may be also called global, as the method of Θ estimation is global.

In the second step the nonparametrical estimation of residuum $\tilde{r}(x)$ is performed:

$$\tilde{f}(x) = \tilde{f}(x, \tilde{\Theta}) \dot{+} \tilde{r}(x)$$

We may see that the parametrical and nonparametrical steps may be joined as addition or multiplication.

3 Periodic Kernels

Periodic kernels belong to the wide group of kernel functions. Besides the standard criterions periodic kernel should fulfil some other conditions. If we assume that the parameter T means the period of the analyzed time series then for every $k \in \mathbb{Z}$ $K(kT)$ should be the strong local maximum. Periodic kernel should also be the connection of two components: periodic, that gives the periodic local maxima and the nonnegative that gives the global maximum at $x = 0$. Both functions should be even because their multiplication will also be even. In this section two periodic kernel are defined.

3.1 First Periodic Kernel (FPK)

One of the simplest form of periodic kernel is as follows:

$$K(x) = \frac{1}{C} e^{-a|x|} (1 + \cos bx) \quad (5)$$

where the C constant assures that the one of the most important kernel function condition is fulfilled: $\int_{-\infty}^{\infty} K(x) dx = 1$. The C value depends on a and b constants:

$$C = 2 \int_0^{\infty} e^{-ax} (1 + \cos bx) dx = \frac{4a^2 + 2b^2}{a(a^2 + b^2)}$$

The form of periodic kernel given with the equation 5 is easy to be analysed but is not interpretable. If we assume that the time series period is T then

$$b(T) = \frac{2\pi}{T}$$

If we define the kernel function attenuation θ as the fraction two following local maxima $\theta = K(t+T)/K(t)$ ($K(t)$ is the local maximum) we may calculate the demanded value of the a parameter:

$$\theta = \frac{K(t+T)}{K(t)} \Rightarrow -aT = \ln \theta \Rightarrow a(\theta) = -\frac{\ln \theta}{T}$$

This leads to the more interpretable form of this periodic kernel:

$$K(x) = \frac{1}{C} e^{\frac{\ln \theta}{T}|x|} \left(1 + \cos \frac{2x\pi}{T} \right) \quad (6)$$

with the C constant given as:

$$C = \frac{4T \ln^2 \theta + 4T \pi^2}{-\ln^3 \theta - 4\pi^2 \ln^2 \theta}$$

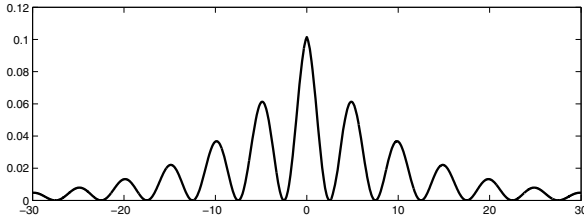


Fig. 1 First Periodic Kernel with $T = 5$, $\theta = 0.6$.

In the equation [6](#) it is easy to point the period of the analysed time series and the impact of the following far distanced observations.

3.2 Second Periodic Kernel (SPK)

First Periodic Kernel is the simplest function that satisfies all conditions of periodic kernel. But this kernel has small disadvantage: its values do not change significantly in the neighborhood of the local maximum. This led to the definition of the Second Periodic Kernel which is more „sharp” at local maxima:

$$K(x) = \frac{1}{C} e^{-a|x|} \cos^n bx \quad (7)$$

with the C constant defined as:

$$C = 2 \int_0^{\infty} e^{-ax} \cos^n bx dx = 2[I_n]_0^{\infty}$$

where I_n denotes:

$$I_n = \int e^{-ax} \cos^n bx dx$$

Second Periodic Kernel also may be presented with the usage of its period and the attenuation:

$$K(x) = \frac{1}{C} e^{-\frac{\ln \theta}{T}|x|} \cos^n \frac{\pi x}{T}, \quad b(T) = \frac{\pi}{T}, \quad a(\theta) = -\frac{\ln \theta}{T} \quad (8)$$

The third parameter of this kernel n may be called sharpness as it describes the kernel variability in the neighborhood of the local maximum.

Because the integrand function I_n is quite complicated, the general calculations for the value of the C constant are performed for the simplification of SPK, where $b = 1$.

General formula for the indefinite integral of this class of function is given with the recurrent formula as follows:

$$\int e^{ax} \cos^n x dx = \frac{e^{ax} \cos^{n-1} x}{a^2 + n^2} (a \cos x + n \sin x) + \frac{n(n-1)}{a^2 + n^2} \int e^{ax} \cos^{n-2} x dx$$

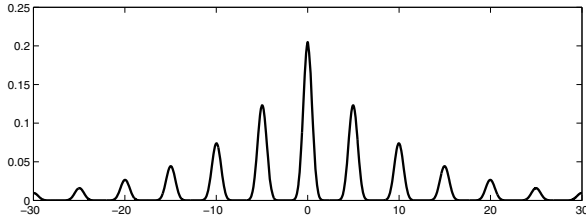


Fig. 2 Second Periodic Kernel with $T = 5$, $\theta = 0.6$ and $n = 10$.

If we define $I_0 = a^{-1} \exp ax$ and $I_n = \delta_n + \mu_n I_{n-1}$ where

$$\mu_i = \frac{2i(2i-1)}{a^2 + 4i^2}, \mu_0 = 1, \quad \delta_i = \frac{e^{ax} \cos^{2i-1} x}{a^2 + 4i^2} (a \cos x + 2i \sin x), \delta_0 = 0$$

then the general formula for I_n is given as:

$$I_n = \left(I_0 + \sum_{i=0}^n \frac{\delta_i}{\prod_{k=0}^i \mu_k} \right) \prod_{i=0}^n \mu_i$$

With $\delta_0 = 0$ and $\mu_0 = 1$ it simplifies into:

$$I_n = \left(I_0 + \sum_{i=1}^n \frac{\delta_i}{\prod_{k=0}^i \mu_k} \right) \prod_{i=1}^n \mu_i$$

To calculate the definite integral we need to find some limiting values of indefinite integral. The general formula of $\lim_{x \rightarrow X} I_n$ is given below:

$$\begin{aligned} \lim_{x \rightarrow X} I_n &= \lim_{x \rightarrow X} \left(I_0 + \sum_{i=1}^n \frac{\delta_i}{\prod_{k=0}^i \mu_k} \right) \prod_{i=1}^n \mu_i = \prod_{i=1}^n \mu_i \lim_{x \rightarrow X} \left(I_0 + \sum_{i=1}^n \frac{\delta_i}{\prod_{k=0}^i \mu_k} \right) = \\ &= \prod_{i=1}^n \mu_i \left(\lim_{x \rightarrow X} I_0 + \sum_{i=1}^n \frac{\lim_{x \rightarrow X} \delta_i}{\prod_{k=0}^i \mu_k} \right) \end{aligned}$$

Four interesting limits take following values:

$$\lim_{x \rightarrow 0^+} C_0 = a^{-1}, \quad \lim_{x \rightarrow \infty} C_0 = 0, \quad \lim_{x \rightarrow 0^+} \delta_i = \frac{a}{a^2 + 4i^2}, \quad \lim_{x \rightarrow \infty} \delta_i = 0$$

So, finally:

$$C = 2[I_n]_0^\infty = \lim_{x \rightarrow \infty} I_n - \lim_{x \rightarrow 0^+} I_n = \left(-\frac{1}{a} - \sum_{i=1}^n \frac{a}{(a^2 + 4i^2) \prod_{k=0}^i \mu_i} \right) \prod_{i=1}^n \mu_i$$

Table 2 Some numerically calculated values of C .

period T	attenuation θ	sharpness n	$C(T, \theta, n)$
1	0.8	2	2.2435
1	0.8	10	1.1056
5	0.8	4	8.4177
5	0.9	4	17.8031
12	0.9	80	20.2726
12	0.4	60	2.8399

In the Table 2 some values of C for different period, attenuation and sharpness values are presented.

4 PerKE Algorithm

4.1 Definition

As it was mentioned in the introduction *PerKE* should be included to the group of semiparametric (two step) methods [1, 7]. It assumes that the initial step (identification of the time series trend) was done and the trend is eliminated. After the elimination of the trend the *PerKE* forecast is performed. It takes the place as the typical kernel estimation of the regression function with the usage of Nadaraya-Watson estimator and one of the periodic kernels as the kernel function. The smoothing parameter is set to 1. Prediction of the residuum part is defined as follows:

$$x_t = \frac{\sum_{i=1}^n x_i K(t-i)}{\sum_{i=1}^n K(t-i)} \quad (9)$$

and is based on the Nadaraya-Watson kernel estimator (equation 3). The *PerKE* forecast is put on the trend forecast and it becomes the final prediction.

4.2 Underestimation

It may occur that the prediction results are not satisfactory. We may define the value called underestimation α that is the ratio of predicted value and the real one: $\alpha = \tilde{x}/x$. Then the improvement of predicted value means the division of predicted value and the underestimation. It is not possible to improve prediction result on the basis of the underestimation of the predicted values. So the underestimation is calculated on the subset of training data. If p is the interesting prediction horizon than we exclude from the train time series the last p observation and treat it as a tune set. On the basis of the limited train set the forecast of the tune time series values is performed and its results are used for the calculating the underestimation. There are three basic models of underestimation calculating. If we have p forecasts and p observation we may use p underestimations: different for every part of the forecast, we may use the mean of underestimation or we may use the underestimation of the closest value (the underestimation of the last value from the tune time series).

Table 3 Comparison of the result for the E series.

method	$RMSE$	$MAPE$	annotations
$SARIMA$	10115,91	20,95 %	$(1, 0, 0)(2, 0, 0)^{12}$
decomp.	9010,87	22,10 %	linear + additive
FPK	19855,28	69,13 %	
SPK	8638,12	20,08 %	$T = 12, \theta = 0.9, n = 80$

Table 4 Comparison of the result for the G series.

method	$RMSE$	$MAPE$	annotations
$SARIMA$	26,95	4,80 %	$(1, 1, 0)(1, 0, 0)^{12}$
decomp.	26,60	4,51 %	exponential + multiplicative
FPK	16,10	3,20 %	
SPK	21,00	3,72 %	$T = 12, \theta = 0.4, n = 60$

5 Experiments and Results

For experiments two real time series were taken. First of them was originally presented in [2] and denoted the number of american airlines passengers (in millions, from 01.1949 to 12.1960). This series is called G and contains 144 observations. The second one (denoted as E) describes the monthly production of heat in the one of the heating plant in the Poland (from 01.1991 to 01.1999).

As the prediction quality measure two evaluations were used: mean percentage absolute error and root mean squared error. Their formulas are presented below:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \tilde{y}_i|}{|y_i|} \quad RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{n}}$$

$PerKE$ algorithm was compared with two very popular methods of time series forecasting: decomposition and $SARIMA$. Two kinds of trend were taken into consideration for the decomposition method: linear and exponential. Also two models of decomposition were considered: additive and multiplicative. For the $SARIMA$ model the seasonal delay has quite intuitive value $s = 12$.

In the first step of the $PerKE$ algorithm elimination of the time series trend was performed on the basis of the decomposition method results. The trend and the model that gave the best result in the decomposition method was used to eliminate the trend from data.

6 Conclusions and Further Works

In this short paper the new algorithm of time series prediction was described ($PerKE$). It applies the kernel regression directly into the time series forecasting. The advantage of this method is that there is no need of data transformation as it take place in earlier works [8, 9]. This algorithm was used to predict values of two

real time series. We may see that *PerKE* may give better results than standard methods of time series prediction: *SARIMA* and the decomposition method. Even if different quality measures (*MAPE*, *RMSE*) are taken into consideration as the method verifiers.

Further works will focus on improvement of the periodic kernel parameter evaluation. In the present form both periodic kernel have some tuning parameters like the attenuation or the sharpness. Also the postprocessing step (division by the underestimation), that may be performed in the several ways, should be determined by the analyzed data.

Acknowledgments. This work was supported by the European Community from the European Social Fund.

References

- [1] Abramson, I.S.: Arbitrariness of the pilot estimator in adaptive kernel methods. *J. of Multivar. Anal.* 12, 562–567 (1982)
- [2] Box, G.E.P., Jenkins, G.M.: *Time series analysis*. PWN, Warsaw (1983)
- [3] Epanechnikov, V.A.: Nonparametric Estimation of a Multivariate Probability Density. *Theory of Probab. and its Appl.* 14, 153–158 (1969)
- [4] Fan, J., Gijbels, I.: Variable Bandwidth and Local Linear Regression Smoothers. *Ann. of Stat.* 20, 2008–2036 (1992)
- [5] Gasser, T., Kneip, A., Kohler, K.: A Flexible and Fast Method for Automatic Smoothing. *J. of Am. Stat. Ass.* 415, 643–652 (1991)
- [6] Gasser, T., Muller, H.G.: Estimating Regression Function and Their Derivatives by the Kernel Method. *Scand. J. of Stat.* 11, 171–185 (1984)
- [7] Hjort, N.L., Glad, I.K.: Nonparametric density estimation with a parametric start. *Ann. of Stat.* 23(3), 882–904 (1995)
- [8] Michalak, M.: Adaptive kernel approach to the time series prediction. *Pattern Anal. and Appl.* (2010), doi:10.1007/s10044-010-0189-3
- [9] Michalak, M.: Time series prediction using new adaptive kernel estimators. *Adv. in Intell. and Soft Comput.* 57, 229–236 (2009)
- [10] Nadaraya, E.A.: On estimating regression. *Theory of Probab. and its Appl.* 9, 141–142 (1964)
- [11] Scott, D.W.: *Multivariate Density Estimation*. In: *Theory, Practice and Visualization*. Wiley & Sons, Chichester (1992)
- [12] Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, Boca Raton (1986)
- [13] Terrell, G.R.: The Maximal Smoothing Principle in Density Estimation. *J. of Am. Stat. Ass.* 410, 470–477 (1990)
- [14] Terrell, G.R., Scott, D.W.: Variable Kernel Density Estimation. *Ann. of Stat.* 20, 1236–1265 (1992)
- [15] Turlach, B.A.: *Bandwidth Selection in Kernel Density Estimation: A Review*. C.O.R.E. and Institut de Statistique, Universite Catholique de Louvain (1993)
- [16] Wand, M.P., Jones, M.C.: *Kernel Smoothing*. Chapman & Hall, Boca Raton (1995)
- [17] Watson, G.S.: Smooth Regression Analysis. *Sankhya - The Indian J. of Stat.* 26, 359–372 (1964)

Unified View of Decision Tree Learning Machines for the Purpose of Meta-learning

Krzysztof Grąbczewski

Abstract. The experience gained from thorough analysis of many decision tree (DT) induction algorithms, has resulted in a unified model for DT construction and reliable testing. The model has been designed and implemented within Intemi – a versatile environment for data mining. Its modular architecture facilitates construction of all the most popular algorithms by combining proper building blocks. Alternative components can be reliably compared by tests in the same environment. This is the start point for a manifold research in the area of DTs, which will bring advanced meta-learning algorithms providing new knowledge about DT induction and optimal DT models for many kinds of data.

Keywords: Decision trees, meta-learning, object oriented design.

1 Introduction

Numerous articles on decision tree (DT) induction have been published so far and new ones appear from time to time. Each proposed algorithm includes a number of solutions, that can be applied in different combinations, not only the ones originally presented.

There is still a need for thorough research on advantages and drawbacks of different approaches. The most popular books on the subject [2, 22] are devoted to particular algorithms and have been published a long time ago. A newer book publication [24] has also undertaken the subject, but the area remains not exhaustively examined. Some comparisons of DT algorithms have also been published, but they are restricted to some split criteria [3, 16] or pruning methods [4, 17, 21].

The major reason of the lack of satisfactory comparative research seems to be the lack of a versatile environment that would make the research really simple.

Krzysztof Grąbczewski

Department of Informatics, Nicolaus Copernicus University, Toruń, Poland

e-mail: <mailto:kg@is.umk.pl>

<http://www.is.umk.pl/~kg>

Some approaches to extensive DT methods libraries have been undertaken (like MLC++ [13], TDDT [23]), but they also suffer from significant limitations (they are restricted to DTs only, so comparison with other methods is not straightforward, they are not ready for deep advanced analysis at the meta-level, etc.).

Recently, a system called Intemi [9, 10] has been designed and implemented. It constitutes a perfect framework for such tasks. Therefore, this framework has been used to implement the unified model of DT algorithms, presented in this article.

The general approach to DTs is the first step towards advanced meta-learning in the area of DTs. With such a tool, it is not difficult to examine the influence of particular elements of the algorithms to the whole model accuracy, and to gain meta knowledge on how to construct learning algorithms to succeed in the analysis of a given dataset.

This article presents a result of the effort related to two branches of science: computational intelligence on one side and object oriented design on the other. Deep analysis of many DT algorithms and their applications brought a general object oriented framework for construction and testing of a variety of learners. Like in most DT algorithms, the focus is put on classification tasks, however any application of the general model is possible by providing alternative components.

Below, section 2 reviews the most popular DT algorithms, section 3 presents the unified model of DT learning machines, section 4 shows which kinds of building blocks compose particular well known algorithms and section 5 signals the advantages of the framework in some applications.

2 Decision Tree Algorithms

The literature provides an abundance of DT algorithms of quite different nature and their different applications. One of the first algorithms to propose hierarchical splits of feature spaces resulting in the form of DTs was ID3 [20]. It dealt with categorical data features only and used information gain measure to determine the splits. The most popular DT classification learners are C4.5 [22] (an extension of ID3) and CART [2]. They split DT nodes with the information gain ratio and the Gini index criterion respectively. They offered some solutions in DT pruning, validation, dealing with missing values in the data etc. Another approach has used the SSV criterion [7, 8] for node splitting.

Other interesting algorithms of slightly different nature are Cal5 [19] and the family of “statistical” DTs: FACT [14], QUEST [15] and CRUISE [12].

All of the methods listed above are capable of building so-called univariate trees, i.e. trees which split nodes on the basis of the values of a single data feature. Some of them (e.g. CART or FACT) were also proposed in multivariate (usually linear) versions.

Most common multivariate DTs perform hierarchical data splits by finding linear combinations of data features and adequate threshold values for the splits. The linear discrimination learning is the fundamental part of each of these algorithms.

The most popular and most successful approaches (apart from CART, FACT and CRUISE) include OC1 [18], LMDT [25], LDT [26], LTree, Qtree and LgTree [5, 6], DT-SE, DT-SEP and DT-SEPIR [11] and the algorithm exploiting the dipolar criterion [1].

3 Unified View of Learning Decision Trees

A thorough analysis of all the algorithms listed above (and also some other less popular approaches) has brought numerous conclusions on their similarities and differences. The conclusions have resulted in the uniform view described below. From the topmost point of view, the tasks related to building decision tree models can be split into two separate groups:

- algorithms of **tree construction**,
- methods of **tree refinement** that are applied on top of different tree construction algorithms, including various techniques of post-pruning and approaches like iterative refiltering.

They are discussed separately in the following subsections.

3.1 Decision Tree Construction

Following the top-down approach of object-oriented analysis and design, we can determine components of the tree construction algorithms:

- **search strategy** that composes the tree node by node,
- **node splitter** i.e. a procedure responsible for splitting the nodes,
- **stop criterion** i.e. the rule that stops the search process,
- **split perspective estimator** i.e. a procedure that defines the order in which the nodes of the tree are split – when using some stop criteria, the order of splitting nodes may be very important, but in most cases it is irrelevant,
- **decision making module** which provides decisions for data items on the basis of the tree,
- optional **data transformations** that prepare the training data at the start of the process or convert the parts of data at particular nodes.

A sketch of dependencies between the modules is presented in figure 1. It can be viewed in terms of classes and interfaces: each box represents an interface and presents a group of classes implementing the interface. Although one could regard UML diagrams as more appropriate form of presentation for the object oriented design of the system, this form is not used here, because it would take much more space and would be less readable. In the figure, the solid arrows represent submodule relation while the dotted ones show the modules used by the search process.

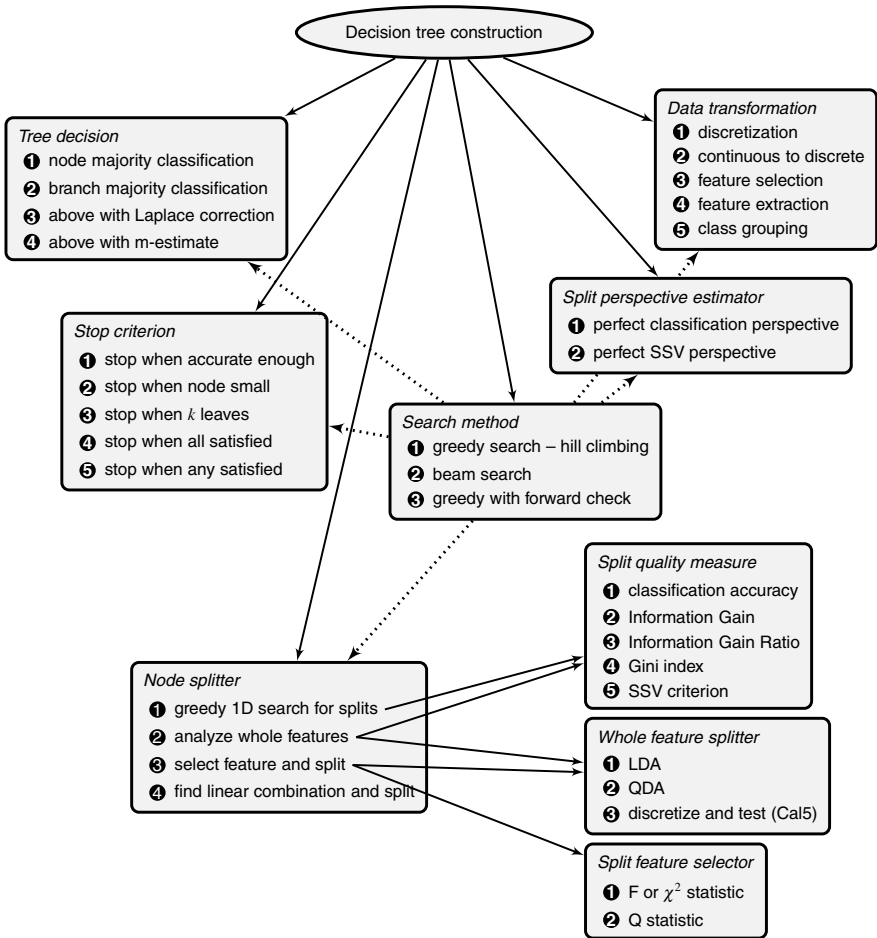


Fig. 1 Information flow between DT construction algorithm modules.

Search Strategies

The search process is the main part of each tree construction algorithm. It uses the functionalities of other modules to grow the tree. Describing shortly: it uses the node splitter to split subsequent tree nodes. The other modules are used to control some details of the process:

- The stop criterion is used to decide when to stop further growth of the tree. Usually, further splits are rejected when the nodes are pure enough or get too small.
- The split perspective estimator rarely affects the resulting tree. It is used to define the order in which the nodes are split. The order may be important when using a stop criterion that acts globally e.g. sends a stop signal when the tree gets a predefined size.

- The data transformation may be used at each node to somehow prepare the node data before the split. It may be a way to implement the technique of LTree family of algorithms, where new features are generated at each node and then analyzed with a split quality measure like information gain. It can also simplify the data at particular node after the split of its parent—for example a feature can be deleted, when it is useless, e.g. all data vectors share a single value of that feature.
- The decision module of the tree does not drive the search process, but the cooperation with the search module may be very advantageous from technical point of view, because the decision module may prepare some information for further decision making just in time, when the information is available i.e. during the search process. When such information is extracted, some information like the training datasets of particular nodes may be discarded from memory, which improves the efficiency of the process.

Almost all approaches to DT induction use the same search method: they split the nodes recursively from the root node until the leaves are obtained. The search process is called a top-down induction, a recursive splits process, a depth first search, hill climbing etc. All the terms, in this context refer to the same algorithm.

The SSV approach and some others performed some experiments with other search methods like beam search or the hill climbing augmented by some insight into further split possibilities and results, but in fact, a thorough analysis of such approaches with reliable conclusions should still be done.

Node Splitters

In the algorithms listed in section 2, three different approaches to node splitting may be observed. Some methods perform a greedy search through all possible splits (sometimes only binary, sometimes also with more than two subnodes): CART, C4.5, SSV, LMDT, OC1 etc. Some others (FACT, QUEST, CRUISE, Cal5) select a feature to split in a separate procedure, and then split according to the selected feature. A version of Cal5 with entropy based split selection, finds a single split for each feature and then uses a split quality measure to select the best one.

Each of the three techniques requires different set of components to be fully configured—a selection of methods from three groups: split quality measures, whole feature splitters and split feature selectors (see the arrows in the figure 1).

Decision Making Modules

The decision making module is usually a classifier that determines target values according to the class distribution in appropriate tree leaf, however some other approaches can also be found. Some researchers try to introduce simple corrections to this decision function (like Laplace correction or m-estimate), but in fact, they do not change the decision, but just soften it, when probabilities of belonging to different classes are provided. In LTree family of algorithms [5], a decision making strategy is proposed that respects class distributions in the nodes of the whole branch of the tree (responsible for classification of particular data item), so the decision may significantly differ from the one taken on the basis of the leaf only.

3.2 *Decision Tree Refinement*

Since in noisy domains DT algorithms are prone to overfit the training data, some techniques of improving generalization abilities had to be worked out. Simple stop criteria have proved to be far from satisfactory and in many publications they are referred to as significantly worse than numerous post-pruning methods. Therefore, most approaches build complete trees, possibly overfitting the training data and then prune them with methods of two groups:

- algorithms based on statistical tests (Pessimistic Error Pruning, Error-Based Pruning, Minimum Error Pruning) or on the Minimum Description Length principle,
- validation based methods.

The techniques belonging to the first group are in general much faster than those of the second group, because they do not test on external data to determine which nodes to prune, but just analyze the training process with statistical tests. Unfortunately, they often result in less accurate trees in comparison to the methods of higher computational cost.

The second group of methods may be further split to the ones that use a single validation dataset (Reduced Error Pruning is the most common technique, but in fact, all the methods of the second subgroup can be applied also in this way) and those, performing multiple training and test cycles (Cost-Complexity Pruning, Critical Value Pruning and degree based pruning of SSV). The last ones are naturally the most time consuming.

The unified framework described here includes all the approaches mentioned above as instances of the two (or three, if those with single validation pass are regarded as separate) general schemes.

3.3 *Other Details*

Some DT algorithms use additional techniques which are not visible in the figure. These include different ways of dealing with missing data or reflecting classification error costs in the tree building processes and in final decision making. Such techniques are easily incorporated into data transformations (missing value imputations), split quality measures (error costs), tree decision making (surrogate splits of CART), etc. The space limitations do not let us go into more detail here.

4 **Well Known Algorithms as Instances of the Uniform Approach**

All the algorithms listed in section 2 perfectly fit the unified view presented in section 3. They have been decomposed into suitable modules. Table 1 presents general information about the types of modules required to realize the goals of particular algorithms. It is not possible to show all the details in such compact visualization.

Table 1 Most popular DT algorithms in terms of the unified model.

	ID3 [20]	C4.5 [22]	CART [2]	SSV [7, 8]	Cal5 [19]	FACT [14]	QUEST [15]	CRUISE [12]	OC1 [18]	LMDT [25]	LDT [26]	LTree family [5, 6]	DT-SE family [11]	Dipolar criterion [1]
Initial data transf.	①					②								
Search method	①	①	①	① ② ③	①	①	①	①	①	①	①	①	①	①
Split perspective m.	①	①	①	① ②	①	①	①	①	①	①	①	①	①	①
Stop criterion	①	①	①	①	①	①	①	①	①	①	①	①	①	①
Node data transf.						②	③							
Node splitter	①	①	①	①	② ③	③	③	③	④	④	④	④	④	④
Split quality m.	②	③	④	① ⑤	②							③		
Whole feature splitter					③	①	②	① ②						
Split feature selector					②	①	①	①						
Tree decision	①	①	①	①	①	①	①	①	①	①	①	②	①	①

Some table cells contain more than one numbered bullet indicating that the methods have several variants. Some options are deliberately omitted to keep the table readable, for example, all but one algorithms are assigned ① in the row of tree decision, while in fact, many algorithms were tried also with Laplace correction or m-estimates of probabilities.

5 Framework Facilities

The preceding section illustrates how different popular DT induction methods can be constructed from components. Such architecture is very beneficial, when an analysis on the meta-level needs to be performed. It is especially useful in meta-learning, which seems to be the future of data mining.

The framework is being successfully used in research activities concerning different aspects of DT induction. It facilitates as just comparisons of different components in action, as possible. For example, to perform a reliable comparative test of different split criteria, we embed each competing component into the same surroundings consisting of a repeated cross-validation of a DT induction process specified by the search method, validation method, decision module etc. Thanks to providing the same environment to all the competing modules we guarantee the same training and test data in corresponding passes, even in the case of inner validation if required by the test scenario. After collecting the results from such test procedures, full information about corresponding results is available, so statistical tests like paired t test, Wilcoxon test or even McNemar test (which requires the information about correspondence between single classification decisions, not only

between the mean accuracies for the whole test datasets), can be applied. In the same way, we compare other types of components like data transformations, stop criteria, validation methods etc.

Conducting the test is quite easy with such framework at hand implemented within as flexible machine learning environment as Intemi.

The information from such tests provides very precious meta-knowledge, to be used in further meta-learning approaches and eventually to compose more accurate DT induction methods.

6 Summary

The unified model of decision trees, presented here, generalizes all the popular approaches to decision tree induction. All types of components have been implemented in Intemi, a general data mining framework designed with special emphasis on meta-learning possibilities [9, 10]. Such implementation opens the gates to advanced research on meta-level analysis of decision tree algorithms and their particular components. Now, we are supplied with a versatile and efficient tool facilitating reliable comparisons of different components by testing them in the same environment (keeping all the remaining components the same). The meta-level analysis of the algorithm will certainly bring many interesting conclusions about particular components advantages and drawbacks, will help eliminate unsuccessful methods and build meta-learners capable of adjusting all the components to given data in an optimal or close to optimal way.

Acknowledgements. The research is supported by the Polish Ministry of Science with a grant for years 2010–2012.

References

- [1] Bobrowski, L., Krętownski, M.: Induction of multivariate decision trees by using dipolar criteria. In: Zighed, D.A., Komorowski, J., Żytkow, J.M. (eds.) PKDD 2000. LNCS (LNAI), vol. 1910, pp. 331–336. Springer, Heidelberg (2000)
- [2] Breiman, L., Friedman, J.H., Olshen, A., Stone, C.J.: Classification and regression trees. Wadsworth, Belmont (1984)
- [3] Buntine, W., Niblett, T.: A further comparison of splitting rules for decision-tree induction. *Machine Learning* 8, 75–85 (1992), 10.1007/BF00994006
- [4] Esposito, F., Malerba, D., Semeraro, G.: A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(5), 476–491 (1997)
- [5] Gama, J.: Probabilistic linear tree. In: ICML 1997: Proceedings of the Fourteenth International Conference on Machine Learning, pp. 134–142. Morgan Kaufmann Publishers Inc., San Francisco (1997)
- [6] Gama, J.: Discriminant trees. In: ICML 1999: Proceedings of the Sixteenth International Conference on Machine Learning, pp. 134–142. Morgan Kaufmann Publishers Inc., San Francisco (1999)

- [7] Grąbczewski, K., Duch, W.: A general purpose separability criterion for classification systems. In: Proceedings of the 4th Conference on Neural Networks and Their Applications, Zakopane, Poland, pp. 203–208 (June 1999)
- [8] Grąbczewski, K., Duch, W.: The Separability of Split Value criterion. In: Proceedings of the 5th Conference on Neural Networks and Their Applications, Zakopane, Poland, June 2000 , pp. 201–208 (2000)
- [9] Grąbczewski, K., Jankowski, N.: Versatile and efficient meta-learning architecture: Knowledge representation and management in computational intelligence. In: IEEE Symposium Series on Computational Intelligence (SSCI 2007), pp. 51–58. IEEE, Los Alamitos (2007)
- [10] Grąbczewski, K., Jankowski, N.: Efficient and friendly environment for computational intelligence. Knowledge-Based Systems, 41p. (2011) (accepted)
- [11] John, G.H.: Robust linear discriminant trees. In: AI & Statistics 1995 [7], pp. 285–291. Springer, Heidelberg (1995)
- [12] Kim, H., Loh, W.Y.: Classification trees with bivariate linear discriminant node models. Journal of Computational and Graphical Statistics 12, 512–530 (2003)
- [13] Kohavi, R., Sommerfield, D., Dougherty, J.: Data mining using MLC++: A machine learning library in C++. In: Tools with Artificial Intelligence, pp. 234–245. IEEE Computer Society Press, Los Alamitos (1996), <http://www.sgi.com/tech/mlc>
- [14] Loh, W.Y., Vanichsetakul, N.: Tree-structured classification via generalized discriminant analysis (with discussion). Journal of the American Statistical Association 83, 715–728 (1988)
- [15] Loh, W.Y., Shih, Y.S.: Split selection methods for classification trees. Statistica Sinica 7, 815–840 (1997)
- [16] Mingers, J.: An empirical comparison of selection measures for decision-tree induction. Machine Learning 3, 319–342 (1989)
- [17] Mingers, J.: An empirical comparison of pruning methods for decision tree induction. Machine Learning 4(2), 227–243 (1989)
- [18] Murthy, S.K., Kasif, S., Salzberg, S.: A system for induction of oblique decision trees. Journal of Artificial Intelligence Research 2, 1–32 (1994)
- [19] Müller, W., Wyszotzki, F.: The decision-tree algorithm CAL5 based on a statistical approach to its splitting algorithm. In: Machine Learning and Statistics: The Interface, pp. 45–65 (1997)
- [20] Quinlan, J.R.: Induction of decision trees. Machine Learning 1, 81–106 (1986)
- [21] Quinlan, J.R.: Simplifying decision trees. Int. J. Man-Mach. Stud. 27(3), 221–234 (1987)
- [22] Quinlan, J.R.: Programs for machine learning (1993)
- [23] Rokach, L., Maimon, O.: Top-down induction of decision trees classifiers – a survey. IEEE Transactions on Systems, Man and Cybernetics: Part C 1(11), 1–12 (2002)
- [24] Rokach, L., Maimon, O.: Data Mining with Decision Trees: Theory and Applications. World Scientific, Singapore (2008)
- [25] Utgoff, P.E., Brodley, C.E.: Linear machine decision trees. Technical Report UM-CS-1991-010, Department of Computer Science, University of Massachusetts (1991)
- [26] Yildiz, O.T., Alpaydin, E.: Linear discriminant trees. International Journal of Pattern Recognition and Artificial Intelligence 19(3), 323–353 (2005)

Designing Structured Sparse Dictionaries for Sparse Representation Modeling

G. Tessitore and R. Prevete

Abstract. Linear approaches to the problem of unsupervised data dimensionality reduction consist in finding a suitable *set of factors*, which is usually called *dictionary*, on the basis of which data can be represented as a linear combination of the dictionary elements. In recent years there have been relevant efforts for searching data representation which are based on *sparse* dictionary elements or a sparse linear combination of the dictionary elements. Here we investigate the possibility to combine the advantages of both sparse dictionary elements and sparse linear combination. Notably, we also impose a *structure* on the dictionary elements. We compare our algorithm with two other different approaches presented in literature which impose either sparse structured dictionary elements or sparse linear combination. These (preliminary) results suggests that our approach presents some promising advantages, in particular a greater possibility of interpreting the data representation.

1 Introduction

Unsupervised dimensionality reduction seems to bring relevant benefits in many contexts which include classification, regression and control problems [12, 13]. Among linear techniques for dimensional reduction Principal Component Analysis (PCA) [7] is widely used. The PCA approach enables one to approximate signals as a linear combination of a restricted number of orthogonal *factors* that most efficiently explain the data variance. Accordingly a signal composed of a large number of variables can be represented as a small number of coefficients of the linear combination of the factors. The orthogonal factors typically involve all original variables. This aspect can cause some difficulties, notably for the interpretation of the signals expressed as linear combination of the orthogonal factors. For example, in the context of the research on human actions [12] can be useful to determine which “pieces” of the action are more relevant than others in classifying or controlling the action itself.

G. Tessitore · R. Prevete

Department of Physical Sciences, University of Naples Federico II, Naples, Italy

e-mail: {{prevete, tessitore}@na.infn.it}

In recent years, there has been a growing interest for identifying alternatives to PCA which find interpretable and more powerful representations of the signals. In these approaches the factors or the coefficients of the linear combination are obtained using prior information represented by penalization terms or constraints in a minimization problem. One can isolate two different types of approaches: a) *Sparse factors*. In this case each factor involves just a small number of the original variables (for example see Sparse-PCA [14] and Structured-Sparse-PCA [9]). b) *Sparse linear combination of the factors*. Here an *overcomplete* set of factors is chosen in advance or learned from the data, but the approximation of each signal involves only a restricted number of factors. Hence, the signals are represented by sparse linear combinations of the factors (for example see K-SVD [1] and MOD [6]). In both approaches, the orthogonality constraint on the factors is usually violated, the set of factors is called *dictionary*, and each factor is called *dictionary element* or *atom*.

Here we investigate the possibility of combining the advantages of both approaches simultaneously finding sparse dictionary elements and sparse linear combination of the dictionary elements. Importantly, in addition to sparse dictionary elements we also impose a *structure* on the atoms [9]. To this aim, we propose a method for *Sparse Representation with Structured Sparse Dictionary (SR-SSD)*. Our approach is tested by using both synthetic and real datasets. The rest of the paper is organized as follows: In Section 2, we give the theoretical formulation of the problem. The description of our algorithm and its relation to previously proposed method is presented in Section 3. The results of our algorithm on synthetic and real data, compared with the results of other two standard algorithms, are presented in Section 4. Finally, in Section 5 we discuss the results obtained and draw our main conclusions.

Notations: Bold uppercase letters refer to matrices, e.g., \mathbf{X}, \mathbf{V} , and bold lowercase letters designate vectors, e.g., \mathbf{x}, \mathbf{v} . We denote by \mathbf{X}_i and \mathbf{X}^j the i -th row and the j -th column of a matrix \mathbf{X} , respectively. While we use the notation x_i and v_{ij} to refer to the i -th element of the vector \mathbf{x} and the element in the i -th row and the j -th column of the matrix \mathbf{V} , respectively. Given $\mathbf{x} \in \mathbb{R}^p$ and $q \in [1, \infty)$, we use the notation $\|\mathbf{x}\|_q$ to refer to its l_q -norm defined as $\|\mathbf{x}\|_q = (\sum_{i=1}^p |x_i|^q)^{1/q}$. Given two vectors \mathbf{x} and \mathbf{y} in \mathbb{R}^p , we denote by $\mathbf{x} \circ \mathbf{y} = (x_1 y_1, x_2 y_2, \dots, x_p y_p)^T \in \mathbb{R}^p$ the element-wise product of \mathbf{x} and \mathbf{y} .

2 Problem Statement

Let us define a matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ of n rows in \mathbb{R}^p , each one corresponding to an experimental observation, the learning problem we are addressing can be solved by finding a matrix $\mathbf{V} \in \mathbb{R}^{p \times r}$ such that each row of \mathbf{X} can be approximated by a linear combination of the r columns of \mathbf{V} . \mathbf{V} is the *dictionary*, and the r columns \mathbf{V}^k of \mathbf{V} are the *dictionary elements* or *atoms*. Let us call $\mathbf{U} \in \mathbb{R}^{n \times r}$ the matrix of the linear combination coefficients, i.e., the i -th row of \mathbf{U} corresponds to the r coefficients of the linear combination of the r columns of \mathbf{V} in order to approximate the i -th row of \mathbf{X} . Consequently, \mathbf{UV}^T is an approximation of \mathbf{X} . The learning problem can be expressed as:

$$\min_{\mathbf{U}, \mathbf{V}} \frac{1}{2np} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 + \lambda \sum_{k=1}^r \Omega_v(\mathbf{V}^k) \text{ s.t. } \forall j, \Omega_u(\mathbf{U}_j) < T_0 \quad (1)$$

where $\Omega_v(\mathbf{V}^k)$ and $\Omega_u(\mathbf{U}_j)$ are some norms or quasi-norms that constrain or regularize the solutions of the minimization problem, and $\lambda \geq 0$ and $T_0 > 0$ are parameters that controls to which extension the dictionary and the coefficients are regularized. If one assumes that both regularizations Ω_u and Ω_v are convex, for \mathbf{V} fixed the problem (1) is convex w.r.t. \mathbf{U} and vice versa.

As we want to induce a sparsity for the linear combination of the dictionary elements, we can choose either ℓ_0 or ℓ_1 norm for the coefficients. These norms penalize linear combinations containing many coefficients different from zero. If not specified differently, we make the following choice: $\Omega_u(\mathbf{U}_h) = \|\mathbf{U}_h\|_0$.

Following [9] the structured sparsity of the dictionary elements can be imposed by choosing

$$\Omega_v(\mathbf{V}^k) = \left\{ \sum_{i=1}^s \|\mathbf{d}^i \circ \mathbf{V}^k\|_2^\alpha \right\}^{\frac{1}{\alpha}} \quad (2)$$

where $\alpha \in (0, 1)$, and each \mathbf{d}^i is a p -dimensional vector satisfying the condition $d_j^i \geq 0$, with $i = 1, 2, \dots, s$. The s vectors \mathbf{d}^i allow to define the structure of the dictionary elements. More specifically each \mathbf{d}^i individuates a group of variables corresponding to the set $G^i = \{j \in \{1, \dots, p\} : d_j^i = 0\}$. The norm $\|\mathbf{d}^i \circ \mathbf{V}^k\|_2^\alpha$ penalizes the variables for which $d_j^i > 0$ and therefore induces non zero values for variables v_{jk} with $j \in G^i$. The resulting set of selected variables depends on the contribution of each \mathbf{d}^i as described in [8]. For example, if the vectors \mathbf{d}^i induce a partition on the set $\{1, \dots, p\}$, then the penalization term (2) favours the formation of dictionary elements \mathbf{V}^k composed of non-zero variables belonging to just one part of the partition. From [9], who in turn follows [10], the problem (1) considering (2) can be reformulated as follows:

$$\min_{\mathbf{U}, \mathbf{V}, \mathbf{H}} \frac{1}{2np} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 + \frac{\lambda}{2} \sum_{k=1}^r \left[(\mathbf{V}^k)^T \text{Diag}(\mathbf{Z}^k)^{-1} \mathbf{V}^k + \|\mathbf{H}_k\|_\beta \right] \quad (3)$$

s.t. $\forall j, \|\mathbf{U}_j\|_0 \leq T_0$

where $\mathbf{H} \in \mathbb{R}_+^{r \times s}$ is a matrix satisfying the condition $h_{ki} \geq 0$, and $\beta = \frac{\alpha}{2-\alpha}$.

The matrix $\mathbf{Z} \in \mathbb{R}^{p \times r}$ is defined as $z_{jk} = \left\{ \sum_{i=1}^s (d_j^i)^2 (h_{ki})^{-1} \right\}^{-1}$. Notice that minimizer of (3) for fixed both \mathbf{U} and \mathbf{V} is given in a closed form, and it is equal to $h_{ki} = \bar{h}_{ki} = |y_j^k|^{2-\alpha} \|\mathbf{y}^k\|_\alpha^{\alpha-1}$, for $k = 1, 2, \dots, r$ and $i = 1, 2, \dots, s$, where each $\mathbf{y}^k \in \mathbb{R}^{1 \times s}$ is the vector $\mathbf{y}^k = (\|\mathbf{d}^1 \circ \mathbf{V}^k\|_2, \|\mathbf{d}^2 \circ \mathbf{V}^k\|_2, \dots, \|\mathbf{d}^s \circ \mathbf{V}^k\|_2)$.

In order to solve the problem (3), we follow the usual approach of finding the minimum by alternating optimizations with respect to the values \mathbf{H} , to the coefficients \mathbf{U} and to the dictionary \mathbf{V} . Most methods are based on this alternating scheme of optimization [2].

3 SR-SSD Algorithm

SR-SSD algorithm proposed here is composed of three alternate stages: *Update of the matrix \mathbf{H}* , *Sparse Coding Stage*, and *Structured Dictionary Stage*. Notice that the problem (3) is convex in \mathbf{U} for fixed \mathbf{V} and vice versa.

Update of matrix \mathbf{H} . In this stage, we assume that both \mathbf{U} and \mathbf{V} are fixed and update the \mathbf{H} 's values. As said above, one can update \mathbf{H} by a straightforward equation $h_{ki} = \bar{h}_{ki} = |y_i^k|^{2-\alpha} \|\mathbf{y}^k\|_\alpha^{\alpha-1}$, however in order to avoid numerical instability near zero a smoothed update is used as follows: $\mathbf{H}_k \leftarrow \max\{\bar{\mathbf{H}}_k, \varepsilon\}$ with $\varepsilon \ll 1$.

Sparse Coding Stage. The second stage of the algorithm proposed here consists in updating the \mathbf{U} 's values in such a way to obtain a sparse representation of the signals \mathbf{X} , for fixed both \mathbf{V} and \mathbf{H} . Note that the equation (3) is composed of two terms to be minimized, and the second term does not depend on \mathbf{U} . Therefore, the optimization problem posed in (3) can be, in this stage, reformulated as follows: $\min_{\mathbf{U}} \|\mathbf{X} - \mathbf{UV}^T\|_F^2$ s.t. $\forall j, \|\mathbf{U}_j\|_0 \leq T_0$. There are a number of well-known ‘‘pursuit algorithms’’ which finds an approximate solution for this type of problem (see for example Basis Pursuit (BP) (4) and Orthogonal Matching Pursuit (OMP) (11)). In our approach we use OMP in experiments in Section 4.1 whereas Iterative Soft-Thresholding (IST) algorithm (3) for experiments described in Section 4.2 where we replaced the ℓ_0 with the ℓ_1 norm.

Structured Dictionary Element Stage. The update of the dictionary \mathbf{V} is performed in this stage, and, more importantly, following the approach suggested by (9) a structured sparse representation for the atoms is found. Fixed both \mathbf{U} and \mathbf{H} the problem (3) can be reformulated as follows

$$\min_{\mathbf{V}} \frac{1}{2} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 + \frac{\lambda np}{2} \sum_{k=1}^r (\mathbf{v}^k)^T \text{Diag}(\mathbf{Z}^k)^{-1} \mathbf{v}^k \quad (4)$$

Although in this case both the two terms of the problem (4) are convex and differentiable with respect to \mathbf{V} , for fixed both \mathbf{U} and \mathbf{H} , leading to a closed form solution for each row of \mathbf{V} , in order to avoid p matrix inversions, we consider a proximal method to update \mathbf{V} :

$$\mathbf{V}^k \leftarrow \text{Diag}(\mathbf{Z}^k) \text{Diag}(\|\mathbf{U}^k\|_2^2 \mathbf{Z}^k + np\lambda \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{U}^k - \mathbf{V} \mathbf{U}^T \mathbf{U}^k + \|\mathbf{U}^k\|_2^2 \mathbf{V}^k) \quad (5)$$

where the update rule is obtained by composing a forward gradient descent step on the first term with the proximity operator of the second term of (4).

A description of our method is given in Algorithm 1.

4 Experiments

Two different kinds of experiments were conducted: the first series of experiments is aimed at testing the ability of the proposed method in retrieving the original dictionary from synthetic data. In the second kind of experiments, we test the ability

Algorithm 1 SR-SSD Algorithm**Input:** \mathbf{X} , **Output:** \mathbf{U} , \mathbf{V} **while** stop-criterion is not reached- update \mathbf{H} : closed form solution given by $\mathbf{H}_k \leftarrow \max\{\tilde{\mathbf{H}}_k, \varepsilon\}$ with $\varepsilon \ll 1$.- sparse coding stage: use OMP or IST algorithm to update \mathbf{U}

- dictionary update:

for $k \leftarrow 1$ **to** r

$$\mathbf{V}^k \leftarrow \text{Diag}(\mathbf{Z}^k) \text{Diag}(\|\mathbf{U}^k\|_2^2 \mathbf{Z}^k + np\lambda \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{U}^k - \mathbf{V} \mathbf{U}^T \mathbf{U}^k + \|\mathbf{U}^k\|_2^2 \mathbf{V}^k)$$

endfor**endwhile**

of our approach in finding a meaningful sparse representation of grasping actions which were recorded by means of a dataglove.

4.1 Test 1: Retrieving the Original Dictionary

In this first series of experiments all datasets have been generated by a linear combination of atoms with a fixed structure, and varying the sparsity of both the atoms and the coefficients of the linear combinations. Three methods have been compared: K-SVD [11], which is able to find a sparse representation of the signals with neither structured nor sparse atoms, SSPCA [9] which is able to create structured sparse atoms with no sparse representation of the signals, and the proposed approach SR-SSD.

Experimental Set-up and Procedures. In this series of experiments the i -th signal \mathbf{x} of the dataset is computed as $\mathbf{x} = \mathbf{U}_i \mathbf{V}^T + \varepsilon$, where each row \mathbf{U}_i is sparse, and ε is a noise vector drawn from a Gaussian distribution with zero mean and varying the variance according to the values in table 1. The non zero coefficients of \mathbf{U}_i are again drawn from a Gaussian distribution with zero mean and unit variance. The indexes of the non zero coefficients were chosen in a random way according to a uniform distribution. The number of non zero coefficients of \mathbf{U}_i was varied as reported in table 1.

Following [9] we have organized the elements of \mathbf{V} on a $N \times N$ dimensional grid with $N = p^{\frac{1}{2}}$. Only a fraction of the elements of \mathbf{V} are non zero and correspond to a square on the grid. The number of non zero elements was chosen in order to have a certain degree of sparsity for the atoms as reported in table 1. Thus, for a given degree of sparsity, the size of the square is fixed while its position is chosen in a random way. The vectors \mathbf{d}^i are chosen in such a way to favour atoms with non zero elements corresponding to a square on the grid as suggested in [9]. In order to compare the computed dictionary with the original dictionary we have used the same procedure proposed in [11]. In particular for each atom of the original dictionary \mathbf{V} we search the closest column in the computed dictionary $\tilde{\mathbf{V}}$ where the distance between two atoms \mathbf{V}^j and $\tilde{\mathbf{V}}^k$ is defined as $1 - |(\mathbf{V}^j)^T \tilde{\mathbf{V}}^k|$. A distance less than 0.01 is considered a success. For both SSPCA and SR-SSD methods the regularization

Table 1 Parameters used for Test 1.

(r) num. atoms	50
(p) dim. signals	400
(L) num. atoms for each signals (SR-SSD and K-SVD only)	2, 8, 16
(ε) noise std σ	0, 0.025, 0.05, 0.075, 0.125
(n) number of signals	250
percentage of zero elements of the atoms	75%, 90%, 99%
$\log_2(\lambda)$ is searched in the range	$[-11, -19]$ at step -1

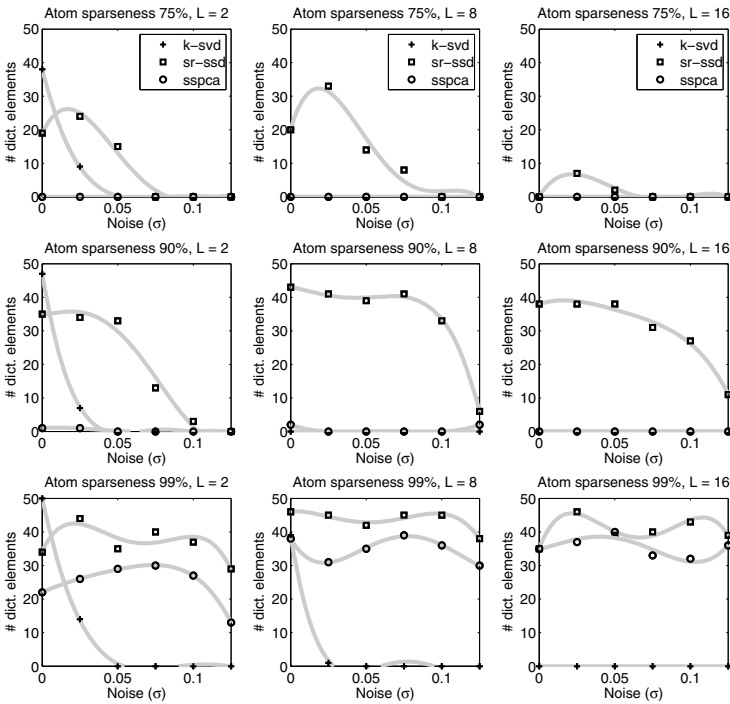


Fig. 1 Performance in retrieving the original structured dictionary. The figure shows the performances of K-SVD, SR-SSD, and SSPCA in retrieving the original dictionary, composed of 50 atoms, against the variation of the atom sparseness (on each row), the coefficient sparseness (on each column), and the noise level (in each plot).

parameter λ has been chosen by a 5-fold cross validation on the representation error. For all methods the number of atoms to retrieve has been set to the number r of atoms composing the original dictionary.

Results. In Figure 1 the results of Test 1 were reported. As expected K-SVD algorithm gives good performance when there is a strong sparsity of the coefficients (see first column). However, the performance decreases rapidly as soon as the noise

variance or the sparsity of the atoms increases. On the other hand, the SSPCA algorithm gives a good performance when sparsity of the atoms is very high (see last row) and the sparsity of the coefficients is not so high (compare first and third column in the last row). SR-SSD algorithm performs very well when there is sparsity on both coefficients and atoms (see last two rows) and it seems more tolerant to noise with respect to both K-SVD and SSPCA. Moreover a relative good performance for SR-SSD is also retained when the sparsity on the atoms is not so high (see first row).

4.2 Test 2: Finding a Meaningful Decomposition of Grasping Actions

Here we compare our method SR-SSD with SSPCA method. The experiments are aimed at testing the ability of both methods in finding a “compact” representation of grasping actions executed by human beings preserving the possibility to identify meaningful parts of the actions which are most discriminant in a classification process.

Experimental Set-up and Procedures. The dataset was composed of two different types of grasps: *precision-grasp* and *power-grasp*. In performing a power-grasp, the object is held in a clamp formed by fingers and palm; in performing a precision-grasp, the object is pinched between the tip of index finger and the opposing thumb. A *bottle top* and a *tennis ball* were used for the precision-grasp and power-grasp, respectively. The actions were recorded by means of the HumanGlove (Humanware S.r.l., Pontedera (Pisa), Italy) endowed with 16 sensors.

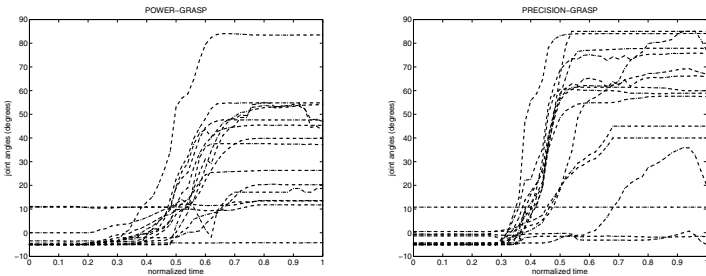


Fig. 2 Example of (left) precision-grasp and (right) power-grasp action recorded by means of the dataglove. Each graph shows the temporal profile of the 16 hand joint angles.

A total of 40 grasping actions were recorded (20 for each type) executed by one subject. A subject was seated at a table with two clearly visible surface marks (m_1 and m_2) placed at a distance of roughly 40 cm from each other. For each target object, the subject was asked to position the right hand on starting position m_1 and in a prone position, and to reach and grasp the target object placed on mark m_2 . All actions were aligned to a fixed length Len and a generic input vector \mathbf{x} was construed as the concatenation of all sensors for all times thus having a dimension of $Len \times 16$.

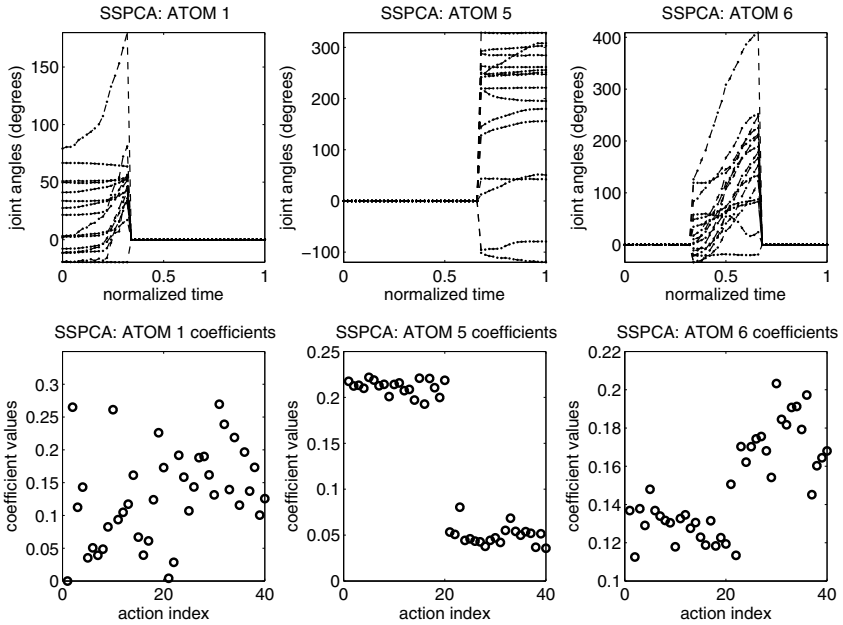


Fig. 3 Dictionary computed by SSPCA. The figure shows (top) 3 out of 6 atoms computed by SSPCA together with (bottom) their coefficients.

A simple dictionary structure was imposed in order to select variables related to the first, the central, and the latter part of the action. To this aim three vectors \mathbf{d}^i were used with elements $d_j^i = 1$ if $\frac{(i-1)*Len}{3} + 1 \leq j \leq \frac{i*Len}{3}$ otherwise $d_j^i = 0$. For both methods six atoms suffice to have a reconstruction error of the data ≤ 0.3 . As for test 1, cross validation was used to find the penalization parameters λ for both methods. For SR-SSD since we do not want to impose a fixed number of coefficients for each signal we prefer to work with an ℓ_1 IST algorithm [5].

Results. In Figures 3 and 4 are reported 3 out of 6 computed atoms together with the corresponding coefficients. The actions are ordered so that the first 20 coefficients are related to precision-grasp actions and the remaining 20 to power-grasp actions. As one can see from Figure 3, SSPCA method tends to select atoms with elements $\neq 0$ corresponding to just one of three consecutive parts of the actions as imposed by the vectors \mathbf{d}^i . In particular, the coefficients of the first atom (see Figure 3) are almost all different from zeros with similar values for both action classes. This is not surprising since the first part of the action is the less informative in discriminating the two classes as the hand always starts in the same prone position. On the other hand the proposed SR-SSD method selects atoms with elements > 0 corresponding to the second and the third part of the actions only, which are the action parts most informative in terms of discriminating the two classes of action (see Figure 4).

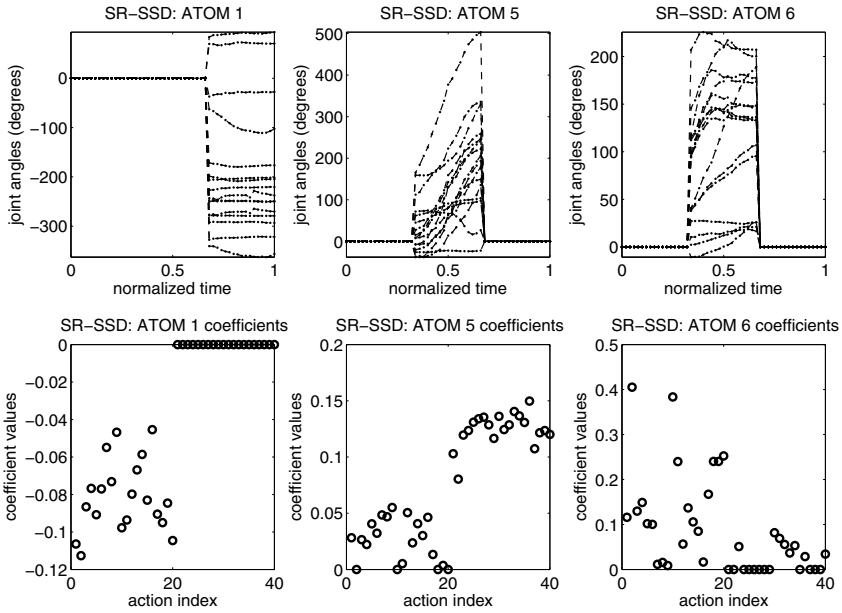


Fig. 4 Dictionary computed by SR-SSD. The figure shows (top) 3 out of 6 atoms computed by SR-SSD together with (bottom) their coefficients.

5 Conclusions

In this paper, we proposed a method to obtain a sparse data representation using a structured sparse dictionary. The results of experiments in Section 4.1 show that the proposed approach (SR-SSD) has the benefits of both K-SVD and SSPCA which use either sparse data representation or structured sparse dictionary. Moreover, results of experiments in Section 4.2 show how the proposed algorithm could be useful for obtaining an interpretable data dimensionality reduction. To this regard note that one could correctly classify the actions on the basis of SSPCA or SR-SSD coefficients. However, the proposed method is also able to select only those “parts” of the actions that most probably contribute in positive to the classification of the actions.

Acknowledgements. This work was partly supported by the project Dexmart (contract n. ICT-216293) funded by the EC under the VII Framework Programme and from the project Action Representations and their Impairment (2010-2012) funded by Fondazione San Paolo (Torino) under the Neuroscience Programme.

References

- [1] Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* 54(11), 4311–4322 (2006)

- [2] Basso, C., Santoro, M., Verri, A., Villa, S.: Paddle: Proximal algorithm for dual dictionaries learning (2010), CoRR, abs/1011.3728
- [3] Bredies, K., Lorenz, D.: Linear Convergence of Iterative Soft-Thresholding. *J. Fourier Anal. Appl.* 14(5-6), 813–837 (2008)
- [4] Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing* 20(1), 33–61 (1998)
- [5] Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics* 57(11), 1413–1457 (2004)
- [6] Egan, K., Aase, S.O., Hakon Husoy, J.: Method of optimal directions for frame design. In: *Proc. of ICASSP 1999*, vol. 5, pp. 2443–2446. IEEE Computer Society, Los Alamitos (1999)
- [7] Hastie, T., Tibshirani, R., Friedman, J.H.: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, corrected edition. Springer, Heidelberg (2003)
- [8] Jenatton, R., Audibert, J.Y., Bach, F.: Structured variable selection with sparsity-inducing norms. Technical report, arXiv:0904.3523 (2009)
- [9] Jenatton, R., Obozinski, G., Bach, F.: Structured sparse principal component analysis. In: *International Conference on AISTATS* (2010)
- [10] Micchelli, C.A., Pontil, M.: Learning the kernel function via regularization. *Journal of Machine Learning Research* 6, 1099–1125 (2005)
- [11] Tropp, J.A.: Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory* 50, 2231–2242 (2004)
- [12] Vinjamuri, R., Lee, H.N., Mao, Z.H.: Dimensionality reduction in control and coordination of the human hand. *IEEE Trans. Biomed. Eng.* 57(2), 284–295 (2010)
- [13] Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T.S., Yan, S.: Sparse Representation for Computer Vision and Pattern Recognition. *Proceedings of the IEEE* 98(6), 1031–1044 (2010)
- [14] Zou, H., Hastie, T., Tibshirani, R.: Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics* 15 (2004)

Run-Time Performance Analysis of the Mixture of Experts Model

Giuliano Armano and Nima Hatami

Abstract. The Mixture of Experts (ME) model is one of the most popular ensemble methods used in pattern recognition and machine learning. Despite many studies on the theory and application of the ME model, to our knowledge, its training, testing, and evaluation costs have not been investigated yet. After analyzing the ME model in terms of number of required floating point operations, this paper makes an experimental comparison between the ME model and the recently proposed Mixture of Random Prototype Experts. Experiments have been performed on selected datasets from the UCI machine learning repository. Experimental results confirm the expected behavior of the two ME models, while highlighting that the latter performs better in terms of accuracy and run-time performance.

1 Introduction

In pattern recognition, the accuracy of classifier systems is usually the main concern. However, in real applications, their run-time performance may play an important role as well. In practice, many well-performing classifiers cannot be used in real applications due to the amount of computational resources required for training, testing, or online evaluation¹. Most real-world classification problems are too complicated for a single classifier to solve. The *divide-and-conquer* strategy,

Giuliano Armano · Nima Hatami
DIEE - Department of Electrical and Electronic Engineering
University of Cagliari
Piazza d'Armi, I-09123 Cagliari, Italy
<http://iasc.diee.unica.it>

¹ We assume that a classifier system behaves in the same manner when testing and online evaluation is performed. Indeed, the difference holds only at an abstract level (i.e., it is related with the interpretation of such phenomena) rather than at a concrete level. For this reason, in the following, testing and online evaluation will be considered in the same way.

e.g., through a combination of classifiers with complementary properties, has proved to be efficient in many of these complex situations. Although the performance problems caused by a large number of classifiers can be coped with using parallel computation, depending on the application and on the available computational power, the designer of a classifier system may be compelled to choose a trade-off between error rate and run-time performance.

The issues to be dealt with upon the adoption of a classifier system that embeds several base classifiers are typically the following: (i) how to divide the problem into simple subproblems and (ii) how to assign base classifiers to solve these subproblems, and (iii) how to obtain the final decision using the outputs of these base classifiers (experts hereinafter).

Jacobs et al. [1,2] have proposed an ensemble method based on the divide-and-conquer strategy, called mixture of experts (ME), in which a set of expert networks is trained together with a gate network. This tight coupling mechanism (i) encourages diversity among experts by automatically specializing them for different regions of the input space and (ii) implements an effective combination policy (i.e., weighted averaging) by dynamically calculating the weights depending on the input in hand.

Recently, Armano and Hatami [6] have proposed a modified ME model, called Mixture of Random Prototype-based Experts (MRPE), which randomly selects some prototype points from the input space and partitions it according to the nearest distance from these prototypes. This strategy adopts a weighting policy based on distances in both training and testing/online evaluation. Instead of a complex gating network, that requires a training process for adjusting its weight parameters, this strategy reduces the overall number of ME parameters (as the weight parameters are removed from the gating networks), thereby simplifying the search and reducing the time required for training ME experts.

Although the ME model has been deeply investigated, to our knowledge, no remarkable work has been made so far on the analysis of its run-time performance in the training and testing/online evaluation phases. This paper first analyzes the expected run-time performance of the ME model. This can be helpful to compare different classifiers and machine learning algorithms with the ME model, not only in terms of accuracy but also from the point of view of run-time analysis. Then, the ME and the MRPE models are compared according to this perspective. Experiments have been performed on selected datasets from the UCI machine learning repository. Experimental results report the outcome for the two ME models in terms of error rate, expected training time, and testing/evaluation time.

The rest of the paper is organized as follows: in Section 2, we briefly recall the standard ME and the MRPE models. In Section 3, we introduce a metric to perform run-time analysis of both models. Experimental results are reported and discussed in Section 4. Section 5 concludes the paper and briefly outlines future research directions.

2 Mixture of Local Experts

2.1 The Standard ME Model

The standard ME model (adaptive mixture of local experts) [1,2] adopts a learning procedure which achieves improved generalization performance by assigning different subtasks to different experts. Its basic idea consists of concurrently training several experts and a gating network. The gating function assigns an estimate of a probability to each expert, based on the current input. In the training phase, this value denotes the probability for a pattern to appear in the training set. In the testing/online evaluation phase, it defines the relative contribution of each expert to the ensemble. Training is aimed at achieving two goals: (i) find the optimal gating function for each expert and (ii) optimize the accuracy of each expert according to the probability distribution (for an input to “belong” to a given expert) estimated by the gating function. Needless to say that to which extent the above goals are reached depends on the effectiveness of the training phase. In other words, resulting misclassifications in this model derive from two sources: (a) the gating network is unable to correctly estimate the probability for a given input sample and (b) experts did not perfectly learn their subtask.

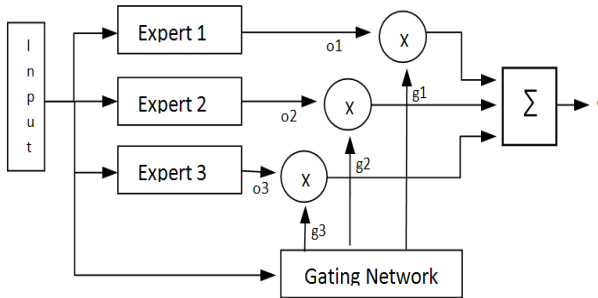


Fig. 1 Block diagram representing the mixture of experts (ME). The generic model shown here has three experts ($N_e = 3$) and the gating network is a mediator for managing the process of combining experts.

Let us now consider the network shown in Fig. 1, which reports a generic ME model with $N_e = 3$ experts. The i th expert produces its output $o_i(x)$ as a generalized linear function of the input x :

$$o_i(x, W_i) = f(W_i \cdot x) \quad (1)$$

where W_i is the weight matrix of the i th expert and $f(\cdot)$ is a predefined continuous nonlinearity. The gating network is also a generalized linear function, and its i th output, $g_i(x, V_i)$, is the multinomial logit (or softmax function) of the gating network output, o_{g_i} .

$$g_i(x, V_i) = \frac{\exp(o_{g_i})}{\sum_{j=1}^{N_e} o_{g_j}} \quad i = 1, \dots, N_e \quad (2)$$

where V_i is the weight vector of the gating network. Hence, the overall output of the ME architecture, $o(x)$, is

$$o(x) = \sum_i g_i(x, V_i) \cdot o_i(x, W_i) \quad (3)$$

Two training procedures are suggested in the literature [1,3] for finding optimal weight parameters W_i and V_i . The first is the standard error back-propagation algorithm with gradient descent, whereas the second is based on the Expectation-Maximization (EM) method.

2.2 Mixture of Random Prototype-Based Experts

The key idea of MRPE is to randomly partition the input space of the problem into subspaces and then let each expert specialize on each subspace by means of “soft” competitive learning. First, the input space is partitioned according to some prototypes randomly chosen from the training set, so that the input samples are weighted during the training and testing phases based on their distances from the selected prototypes. The main advantage of this method is that, instead of a complex gating network which must be trained concurrently with other experts, the proposed gating function has no parameters (weights) to adjust, as it simply enforces a distance-based weighting policy. This modification improves three important aspects of the standard ME model. First, it reduces the training time by decreasing the number of parameters to be estimated. Secondly, as simple distance measures used by the gating function are more robust with respect to errors in determining the area of expertise of an expert, errors in the proposed ME model are mostly due to the error made by the expert networks, with an expected positive impact on the overall accuracy. Lastly, the region of expertise for each expert in the standard ME model is nested, which makes the problem difficult to learn. In the proposed method, the area of expertise of each expert is more centralized, which makes the subproblem easier to learn.

In the case of disjoint (or *hard*) partitioning, a system based on the MRPE model would first measure the distance between each training sample and the prototypes, and then assign a fixed value η_j to the weight h_i of the i -th expert, proportional to these distances. In principle, h_i is expected to be an estimate of the posterior probability for the i th expert to generate the desired output o_i , and is used as the coefficient of the learning rate for updating the weight parameters of the expert. This implies that the weights of the expert network whose prototype is nearest to the current input sample, will be updated more than those belonging to the other experts. Similarly, in the testing phase, the expert whose prototype is nearest to the input sample will contribute to a greater extent to the final output. Unlike disjoint partitioning, where the learning rate coefficients are the same for one partition element and change sharply from one to another, in the overlapping (or *soft* partitioning) method they change

smoothly, proportional to the distances. Similarly, the value of d_i for the i th expert depends on how close the prototype embedded by the expert is to the current input sample. Summarizing, learning a disjoint-partitioning model the amount of expertise and contribution of experts is fixed for each partition element, whereas in the overlapping learning their expertise smoothly vary with the distance between the embedded prototype and the current input. The full detailed algorithm can be found in [6].

3 Run-Time Performance of ME and MRPE

In computer science, the worst-case time complexity of algorithms is typically evaluated in terms of asymptotic behavior, denoted by the Landau symbol O (also called “big- O ” notation) [7]. For practical applications, the asymptotic behavior is not informative enough, as the order of magnitude expressed in terms of well-known functions (e.g., logarithmic, polynomial, and exponential) hides details that can make the difference in terms of elapsed time for a run. For this reason, we prefer to carry out our analysis in terms of number of floating point operations (FLOPs). This is not an ideal measure [8]; however, it is useful for comparing the performance of a classifier system, as it focuses on the number of additions and multiplications required to perform a given task. In modern computer systems with floating point units (FPUs), addition and multiplication are comparable in complexity. Division has a similar complexity and function evaluations (e.g. $\exp(x), x^p$) have roughly 2.5 times the complexity of an addition.

The expected run-time performance of any learned classifier system, including those compliant with the ME model, requires different formulations, depending on which specific phase (i.e., training or testing/online evaluation) is considered. As training is usually performed once (or from time to time in the event that the underlying process is time-variant), in the following we will concentrate on testing/online evaluation.

3.1 Run-Time Performance of the ME Model

In general, this run-time performance depends on the type of classifier used, on its parameters, and on the characteristics of the problem to be solved (e.g., the number of samples used for training n , the number of features (dimensions) in each sample d , and the number of classes to be distinguished c).

As reported in Figure 2 the run-time performance of an ME classifier system can be decomposed as three main components: 1) expert networks, 2) gating network, and 3) aggregation (also called combination).

As clearly shown in Figure 2 the overall testing/online evaluation performance of an ME classifier equipped with N_e experts is:

$$T_{ME} = \sum_{e=1}^{N_e} T_e + T_g + T_a \quad (4)$$

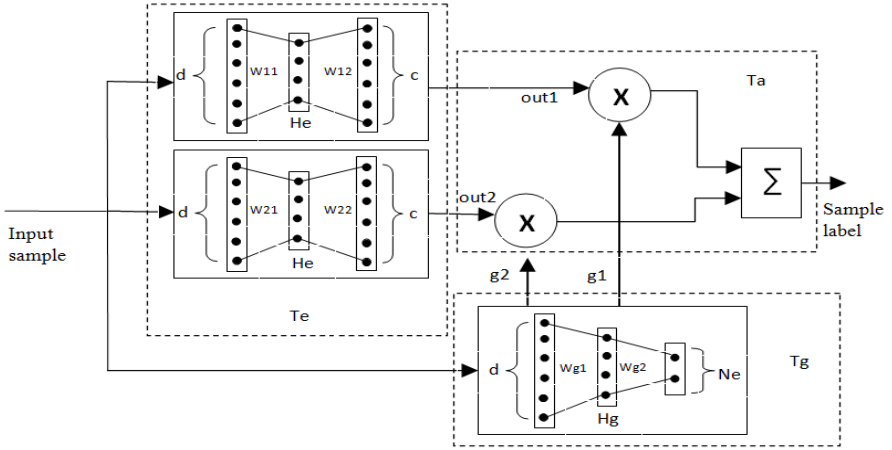


Fig. 2 The expected run-time performance of an ME classifier system should be evaluated according to three main components: 1) expert networks (T_e), 2) gating network (T_g) and 3) aggregation policy (T_a).

1. *Run-time performance of an expert (T_e):* Let us consider an expert with d nodes in input, H_e nodes in the hidden layer and c nodes in the output layer. According to these hypotheses, between the input and the hidden layer, and between the hidden and the output layer, there are $W_{i1} = d \times H_e$ and $W_{i2} = H_e \times c$ connections, which yield to $d \times H_e$ and $H_e \times c$ multiplications, $d \times H_e$ and $H_e \times c$ additions, respectively. Considering $H_e + c$ additions of bias terms and $H_e + c$ evaluations of the unit transfer functions in the output nodes (e.g., sigmoids), the total complexity of an individual expert implemented by a Multi-Layer Perceptron (MLP) and equipped with a single hidden layer is:

$$T_e = 2H_e(d + c) + 3.5(H_e + c) \quad (5)$$

2. *Run-time performance of the gating network (T_g):* In the standard ME model, the technology used for implementing the gating network is typically the same of that used for implementing base classifiers. Hence, let us assume that gating networks are also MLPs. However, parameters of the gating function, such as the number of nodes in the hidden and in the output layer, are usually different, as the task of the gating function is different from the one performed by experts. For instance, an implementation of the ME model equipped with $N_e = 3$ experts and applied to a 10-class problem, the output layer of each expert has 10 neurons while the output layer of the gate has 3 nodes. H_g being the number of hidden nodes, the complexity of the gating network formulates as follows:

$$T_g = 2H_g(d + N_e) + 3.5(H_g + N_e) \quad (6)$$

where N_e is the number of individual experts contributing in the ME classifier.

3. *Run-time performance of aggregation(T_a):* In the aggregation step, the final label for each incoming sample is obtained by summing up, suitably weighted, the outputs of the experts (see Figure 2). As the weighted output of an expert has c multiplications, T_a formulates as follows:

$$T_a = c \cdot (N_e + 1) \quad (7)$$

3.2 Run-Time Performance of the MRPE Model

The experts and the aggregation rule in the MRPE model do not change with respect to the standard ME model. Hence, we only need to reformulate the expected run-time performance for the gating network. Indeed, the main contribution of the MRPE is its simple gate structure which makes it less complex *and* more accurate. The distance-based gating function does not need any training process, as it assigns learning rates and weights to an expert only based on the distance between the incoming samples and the random prototypes. The complexity of this distance measure $D(x, p_i)$, where x and p_i are d -dimensional vectors, is $3d$. At each evaluation, the gating network has to calculate N_e times the distance. Assuming that the function evaluation is $h_i(x) = 1 - \frac{d_i}{\|D(x)\|}$, the expected run-time performance is:

$$T_g = N_e(3d + 2.5) \quad (8)$$

As an example, let us suppose that we have an implementation of the ME model with three experts and a gate network with 20 nodes in its hidden layer. Moreover, let us assume that we must deal with a 20-class problem characterized by a 50-dimensional sample space. The ratio of T_{gME}/T_{gMRPE} is 4.65 which clearly shows a significant saving in terms of evaluation complexity.

4 Experimental Results

To check the validity of the MRPE model, both in terms of expected run-time performance and accuracy, we used some machine learning data sets downloaded from UCI [4]. Table 1 lists all selected data sets, which include real-world and synthetic

Table 1 The main characteristics of the selected UCI datasets.

Problem	# Train	# Test	# Attributes	# Classes
Iris	150	-	4	3
Satimage	4435	2000	36	6
Pendigits	7494	3498	16	10
Letter	20000	-	16	26
Vowel	990	-	11	11
Segment	210	2100	19	7
Glass	214	-	9	7
Yeast	1484	-	8	10

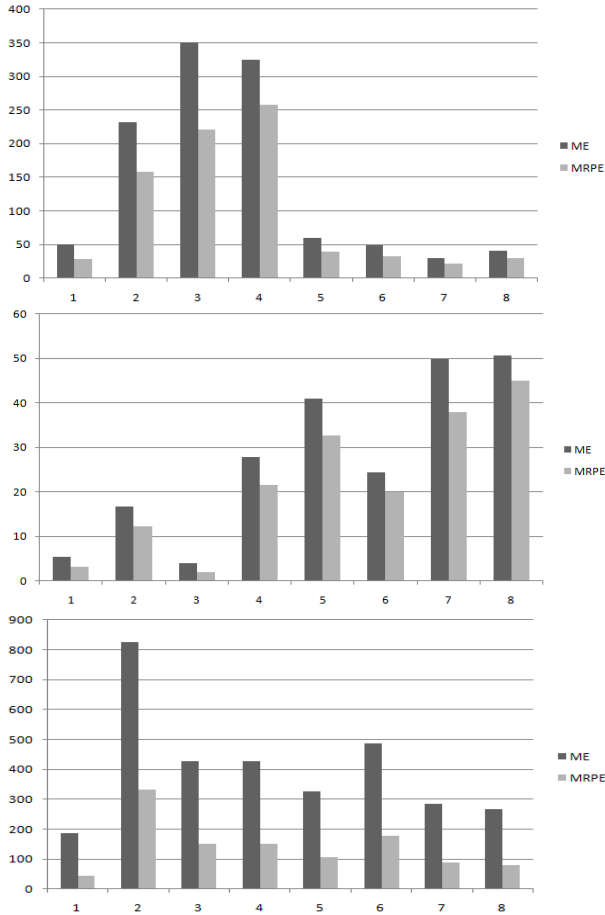


Fig. 3 Performance of the standard ME model vs. the MRPE model, in terms of training time, error rate and testing/online evaluation time, respectively. The numbers in the x-axis represent different datasets (in the order mentioned in Table 1).

problems. We used 10-fold cross-validation to ensure statistical significance while evaluating the accuracy of classifiers. To build the standard ME and the MRPE models, we used a MLP architecture with one hidden layer, trained with the back-propagation learning rule [5]. To determine the best value for the N_e number of partitions, which is equal to the number of experts, we varied it from 2 to 10 for each dataset. We also varied the number of hidden neurons in expert networks to experimentally find the optimal architecture of the MLP experts for each problem. The results of these experiments are reported for the ME and for the MRPE models, in terms of training time, error rate and testing/online evaluation performance. Figure 3 highlights that the latter model outperforms the former for all selected datasets.

5 Conclusions and Future Work

In this paper, the expected run-time performance for the standard ME model and for the MRPE model have been formulated for the first time. Theoretical results are confirmed by experimental results, which also highlight that MRPE performs better also in terms of accuracy. Experimental results have been reported for error rate, training time and testing/online evaluation time.

Future work will be dedicated to investigate whether the error and the performance measures can be used for automatically estimating the optimal number of experts required to solve a given problem. Extending the run-time analysis to the Hierarchical ME model is another issue that will be soon investigated.

References

- [1] Jacobs, R., Jordan, M., Barto, A.: Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks. Tech rep. University of Massachusetts, Amherst, MA (1991)
- [2] Jacobs, R., Jordan, M., Nowlan, S., Hinton, G.: Adaptive mixtures of local experts. *Neural Computation* 3, 79–87 (1991)
- [3] Jordan, M.I., Jacobs, R.A.: Hierarchical mixtures of experts and the EM algorithm. *Neural Comp.* 6, 181–214 (1994)
- [4] Murphy, P.M., Aha, D.W.: UCI Repository of Machine Learning Databases, Dept. of Information and Computer Science, Univ. of California, Irvine (1994)
- [5] Haykin, S.: *Neural Networks: A Comprehensive Foundation*, 2nd edn. Prentice-Hall, Englewood Cliffs (1999)
- [6] Armano, G., Hatami, N.: Mixture of Random Prototype-Based Local Experts. In: Graña Romay, M., Corchado, E., Garcia Sebastian, M.T. (eds.) HAIS 2010. LNCS (LNAI), vol. 6076, pp. 548–556. Springer, Heidelberg (2010)
- [7] Duda, R., Hart, P., Stork, D.: *Pattern classification*, 2nd edn. John Wiley & Sons, New York (2001)
- [8] Hennessy, J., Patterson, D.: *Computer architecture: a quantitative approach*. Morgan Kaufmann, San Mateo (1990)

Fuzzy Clustering Finding Prototypes on Classes Boundary

Michał Jezewski and Jacek Leski

Abstract. In the presented paper, new clustering method based on minimization of a criterion function is proposed. Its goal is to find prototypes placed near the classes boundary. The proposed method may be useful in applications to classification algorithms. The clustering performance was examined using the Ripley dataset, and the results seem to be encouraging.

1 Introduction

Clustering consists in finding groups (clusters) of similar objects in dataset. Members of the same group are more similar to one another than to members of other groups [3], [9]. Clustering represents unsupervised learning methods, because only objects features are taken into consideration. Usually the cluster is characterized by its center (prototype). In case of two-dimensional data, it is possible to present clustering results graphically. In case of many-dimensional it is not possible and results of clustering are presented with a help of two matrices: $(c \times N)$ -dimensional partition matrix \mathbf{U} , which describes memberships of N objects to c clusters, $(t \times c)$ -dimensional prototype matrix \mathbf{V} , which describes location of c prototypes in t -dimensional feature space. Non-fuzzy clustering methods assume that each object belongs to only one cluster. In fuzzy clustering object may partially belong to more than one cluster. Clustering based on minimization of criterion function $J(\mathbf{U}, \mathbf{V})$ is the most popular method for finding the best division from all possible ones.

Clustering plays an important role in many engineering fields, mainly in pattern recognition [4]. In presented paper, new fuzzy clustering method based on minimization of criterion function is presented. The goal of proposed clustering is to find prototypes which are placed near the classes boundary. For this reason, the proposed method may be useful in applications to classification algorithms.

Michał Jezewski · Jacek Leski

Institute of Electronics, Silesian University of Technology,

Akademia 16, 44-100 Gliwice, Poland

e-mail: [michal.jezewski,jacek.leski}@polsl.pl](mailto:{michal.jezewski,jacek.leski}@polsl.pl)

The N -element dataset may be denoted as $\mathbf{X} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$, where y_k is the class label which indicates group assignment of k th object to one of two classes ω_1 ($y_k = +1$) or ω_2 ($y_k = -1$), and $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kt}]^T$ represents the k th object (datum) in t -dimensional feature space. Presented method takes into consideration assignments of object to classes (y_k). Therefore, it may be regarded as a clustering with partial supervision, but in another meaning than the one proposed by Pedrycz [7].

2 New Fuzzy Clustering Method

The following criterion function was proposed

$$J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^m d_{ik}^2 + \alpha \sum_{i=1}^c \left(\sum_{k=1}^N u_{ik} y_k \right)^2, \quad (1)$$

with the constraints

$$\forall_{i=1,2,\dots,c} \sum_{k=1}^N u_{ik} = L, \quad L > 0, \quad (2)$$

where d_{ik} denotes Euclidean distance between the i th prototype and k th object, L is the fuzzy cardinality. According to the assumed constraints, the fuzzy cardinality of each group should be equal to L . The goal of first component of (1) is to create clusters with minimal Euclidean distances between objects and prototypes. The second component of (1) is responsible for creating clusters, which include objects from both classes with similar memberships. Prototypes of such clusters should be placed near the classes boundary. In order to explain the role of second component, memberships of two objects from both classes to one i th cluster will be considered:

1. $u_{i1} = 0.9, y_1 = +1; \quad u_{i2} = 0.3, y_2 = -1,$
2. $u_{i1} = 0.9, y_1 = -1; \quad u_{i2} = 0.3, y_2 = +1.$

From similar membership point of view, both cases (1 and 2) are identical - object from first class belongs to i th cluster with membership three times greater (smaller) than object from second class. Clustering consists in minimization of criterion (1). Using only the sum of products $u_{ik} y_k$, the second case seems to be more favourable, because then the sum is equal to -0.6 , whereas the sum for the first case is $+0.6$. For this reason, the squaring of the sum of products $u_{ik} y_k$ is necessary. In the perfect case, i.e. in case of equal memberships of objects from both classes in each group, the value of the second component of (1) is equal to 0.

The parameter α determines the proportion between both components of criterion (1). The parameter m influences a fuzziness of the clusters - a larger m results in fuzzier clusters. Usually, the $m = 2$ is chosen, and such value was assumed.

The Lagrange multipliers were used for minimization of proposed criterion function. For $m = 2$, the Lagrangian of (1) with constraints from (2) is

$$G(\mathbf{U}, \boldsymbol{\lambda}) = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^2 d_{ik}^2 + \alpha \sum_{i=1}^c \left(\sum_{k=1}^N u_{ik} y_k \right)^2 - \sum_{i=1}^c \lambda_i \left(\sum_{k=1}^N u_{ik} - L \right), \quad (3)$$

where λ_i are Lagrange multipliers, $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_c]^T$. The necessary conditions for minimization of (3) are obtained by differentiating it with respect to $\boldsymbol{\lambda}$ and \mathbf{U} and setting the results equal to 0

$$\forall_{1 \leq s \leq c} \frac{\partial G(\mathbf{U}, \boldsymbol{\lambda})}{\partial \lambda_s} = \sum_{k=1}^N u_{sk} - L = 0, \quad (4)$$

$$\forall_{1 \leq s \leq c} \forall_{1 \leq t \leq N} \frac{\partial G(\mathbf{U}, \boldsymbol{\lambda})}{\partial u_{st}} = 2u_{st} d_{st}^2 + 2\alpha y_t \left(\sum_{k=1}^N u_{sk} y_k \right) - \lambda_s = 0. \quad (5)$$

If we denote

$$u_{sk}^* = \frac{u_{sk}}{\lambda_s}, \quad (6)$$

and divide (5) by λ_s , the following equation for t th object is obtained

$$\forall_{1 \leq t \leq N} 2 u_{st}^* d_{st}^2 + 2\alpha y_t \left(\sum_{k=1}^N u_{sk}^* y_k \right) = 1. \quad (7)$$

It is possible to create similar equation for r th object

$$\forall_{1 \leq r \leq N} 2 u_{sr}^* d_{sr}^2 + 2\alpha y_r \left(\sum_{k=1}^N u_{sk}^* y_k \right) = 1. \quad (8)$$

The subtracting of (7) and (8) and simple transformations leads to

$$u_{sr}^* d_{sr}^2 - u_{st}^* d_{st}^2 = (y_t - y_r) \left(\alpha \sum_{k=1}^N u_{sk}^* y_k \right). \quad (9)$$

The following solution was proposed: On the basis (9), the membership of any t th object to s th cluster (u_{st}^*) is determined by membership of any r th object to this cluster (u_{sr}^*). The formula for u_{st}^* depends on classes, that objects \mathbf{x}_t and \mathbf{x}_r belong to. There are possible three cases:

Case 1. Both objects \mathbf{x}_t and \mathbf{x}_r belong to the same class.

In that case $y_t = y_r = -1$ or $y_t = y_r = +1$. As a result $(y_t - y_r) = 0$, and from the equation (9) the membership of t th object to s th cluster by membership of r th object to this cluster is determined with the formula

$$u_{st}^* = u_{sr}^* \frac{d_{sr}^2}{d_{st}^2}, \quad (10)$$

where $1 \leq s \leq c$ and t, r denote any objects from the same class.

Case 2. Object \mathbf{x}_t belongs to ω_1 class and object \mathbf{x}_r belongs to ω_2 class.

In that case $y_t = +1$ and $y_r = -1$. As a result $(y_t - y_r) = 2$, and the equation (9) takes the form

$$u_{sr}^* d_{sr}^2 - u_{st}^* d_{st}^2 = 2\alpha \sum_{k=1}^N u_{sk}^* y_k. \quad (11)$$

The sum on the right side of (11) includes memberships of all N objects, regardless of classes, that they belong to. If we sort the dataset, i.e. put objects from ω_1 class at the beginning of a list with numbers from 1 to N_1 , then objects from ω_2 class with numbers from $N_1 + 1$ to N , the equation (11) may be written in the form

$$u_{sr}^* d_{sr}^2 - u_{st}^* d_{st}^2 = 2\alpha \sum_{k=1}^{N_1} u_{sk}^* - 2\alpha \sum_{k=N_1+1}^N u_{sk}^*. \quad (12)$$

The object \mathbf{x}_t belongs to ω_1 class, and object \mathbf{x}_r belongs to ω_2 class. Therefore, using the previously established formula (10), equation (12) may be transformed to

$$u_{sr}^* d_{sr}^2 - u_{st}^* d_{st}^2 = 2\alpha \sum_{k=1}^{N_1} u_{st}^* \frac{d_{st}^2}{d_{sk}^2} - 2\alpha \sum_{k=N_1+1}^N u_{sr}^* \frac{d_{sr}^2}{d_{sk}^2}. \quad (13)$$

After transformations we obtain the final equation for the second case

$$\forall_{1 \leq s \leq c} \forall_{\substack{1 \leq t \leq N_1 \\ N_1+1 \leq r \leq N}} u_{st}^* = u_{sr}^* \frac{d_{sr}^2 \left(1 + 2\alpha \sum_{k=N_1+1}^N \frac{1}{d_{sk}^2} \right)}{d_{st}^2 \left(1 + 2\alpha \sum_{k=1}^{N_1} \frac{1}{d_{sk}^2} \right)}. \quad (14)$$

Case 3. Object \mathbf{x}_t belongs to ω_2 class and object \mathbf{x}_r belongs to ω_1 class.

In that case $y_t = -1$ and $y_r = +1$. As a result $(y_t - y_r) = -2$. Doing the same as in previous case, we obtain finally

$$\forall_{1 \leq s \leq c} \forall_{\substack{1 \leq r \leq N_1 \\ N_1+1 \leq t \leq N}} u_{st}^* = u_{sr}^* \frac{d_{sr}^2 \left(1 + 2\alpha \sum_{k=1}^{N_1} \frac{1}{d_{sk}^2} \right)}{d_{st}^2 \left(1 + 2\alpha \sum_{k=N_1+1}^N \frac{1}{d_{sk}^2} \right)}. \quad (15)$$

There are u_{sk}^* memberships in all equations above. Thus, further transformations are necessary. Taking into account the constraint (2) and used substitution (6) we obtain

$$\lambda_s = \frac{L}{\sum_{k=1}^N u_{sk}^*}. \quad (16)$$

Finally, the necessary condition for minimization of (11) with respect to \mathbf{U} may be written as

$$\forall_{1 \leq i \leq c} \forall_{1 \leq k \leq N} u_{ik} = \frac{L u_{ik}^*}{\sum_{k=1}^N u_{ik}^*}. \tag{17}$$

Differentiating (11) with respect to \mathbf{V} and setting the result equal to 0 (necessary condition) leads to known (11) equation determining cluster prototypes

$$\forall_{1 \leq s \leq c} \mathbf{v}_s = \frac{\sum_{k=1}^N u_{sk}^2 \mathbf{x}_k}{\sum_{k=1}^N u_{sk}^2}. \tag{18}$$

In the implementation, the following assumption was made: memberships of all objects to given cluster are determined by membership of the object which is the closest to prototype of this cluster. The value of membership of the closest object was established at 1. In other words, in equations (10), (14) and (15)

$$u_{sr}^* = 1. \tag{19}$$

The algorithm for determining \mathbf{U} and \mathbf{V} can be described as:

0. Initialize prototype matrix $\mathbf{V}^{(0)}$ with random values from the range determined by minimal and maximal values of all dataset dimensions, $k = 1$.
1. Calculate distances from objects to prototypes.
2. If any distance is equal to 0 then stop (owing to division by distances in equations (10), (14), (15)).
3. For each prototype, find the object which is the closest to.
4. Update the fuzzy partition matrix $\mathbf{U}^{(k)}$ using equation (10), (14) or (15), depending on classes, that objects: currently analysed and found in previous point belong to.
5. Transform partition matrix $\mathbf{U}^{(k)}$ according to (17).
6. Update the prototype matrix $\mathbf{V}^{(k)}$ using (18).
7. If $\left| J(\mathbf{U}^{(k+1)}, \mathbf{V}^{(k+1)}) - J(\mathbf{U}^{(k)}, \mathbf{V}^{(k)}) \right| > z$, then $k = k + 1$ and go to 1., else stop.

Remark

Step 2. of above algorithm was introduced as a result of several situations, when prototypes were equal to objects. A simple way to avoid such situations is to start clustering from another initial prototype matrix $\mathbf{V}^{(0)}$.

The established values of clustering parameters are given in Table 1.

3 Results

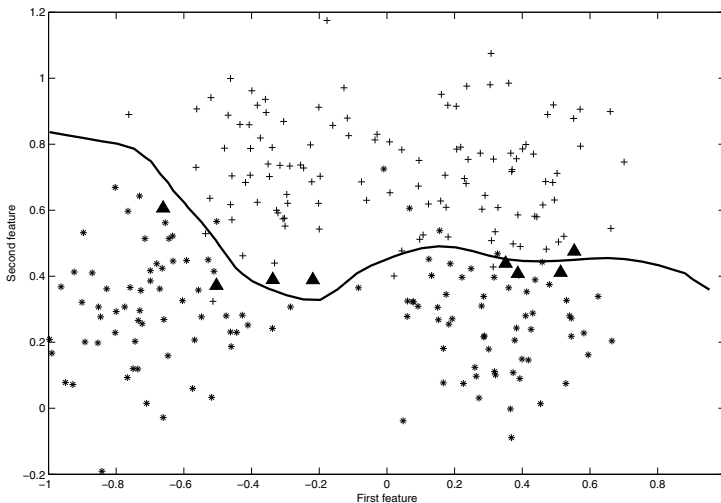
The performance of proposed fuzzy clustering method was examined using the Ripley [8] dataset, which is common benchmark dataset. These data consist of patterns characterized by two features and they are assigned to two classes. Each class has a bimodal distribution obtained as a mixture of two normal distributions. Dataset

Table 1 Values of clustering parameters

Parameter	Value
α	5000
Maximum number of iterations	100
Minimum change of criterion value (z)	10^{-5}
Fuzzy cardinality (L)	10

cardinality is equal to 250 (125 objects from ω_1 class, 125 objects from ω_2 class). According to our other research, instead of original training dataset proposed by Ripley, first training dataset described in [5] was used.

Figures 1 and 2 present two cases of clustering into 8 clusters. Each clustering was performed starting from different initial prototype matrix. Objects from the ω_1 (ω_2) class are marked by star (plus) symbols, respectively. Prototypes obtained after clustering are represented by triangles. The characteristic feature of the proposed method is that sometimes prototypes are placed very close to objects from dataset. Moreover, sometimes one of obtained prototypes is very similar to one another. For example, in figure 2 prototypes number 1 and 3 were very similar, and as a result only 7 triangles are possible to be noticed. In order to obtain expected classes boundary, the LSVM algorithm [6] with the Gaussian kernel was applied. The SVM classifier [2], [10] is regarded as one of the best classification algorithms. The LSVM (Lagrangian SVM) is its modified - faster version. The discrimination curves are represented by black lines.

**Fig. 1** Clustering the Ripley dataset into 8 clusters (case 1)

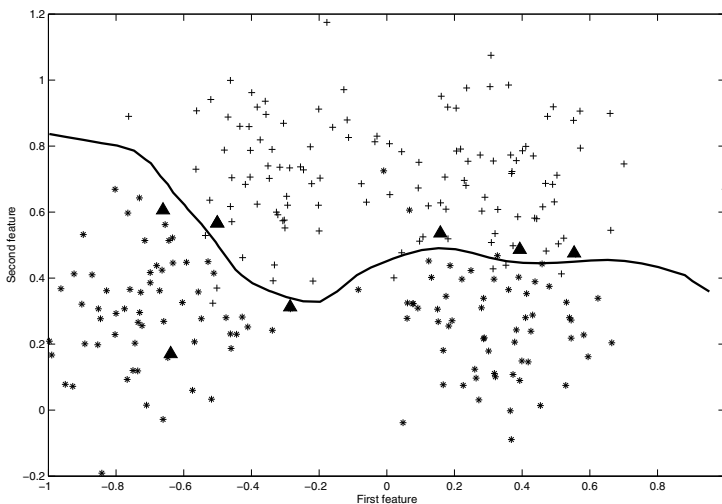


Fig. 2 Clustering the Ripley dataset into 8 clusters (case 2)

After analysis of presented results, in our opinion the proposed method seems to fulfil the assumed idea - prototypes obtained as a result of clustering are placed near the classes boundary. Figure 3 shows clustering into three groups. The black circles represent the initial, randomly chosen prototypes, and the prototype moving process during iterations was marked by black lines and points. This case of clustering also seems to confirm its appropriate working. Firstly, the prototype number 1 is not wrong placed as it may seem, because detailed analysis showed, that membership of a single object from the class ω_1 (single star) was much higher than memberships of the closest objects from the ω_2 class. Secondly we may observe, that if the initial prototype was placed rather good (like prototype number 1), then it moved only a little. From the other hand, if the initial prototype was in wrong place (like numbers 2 and 3), the appropriate moving process was performed. Although the fact, that in this specific case similar memberships of object from both classes were obtained, it is easy to observe, that prototype number 1 should be located in another place. Our further works will concern that problem.

In most of clustering methods, the obtained result depends on initial prototype (partition) matrix. We may observe similar tendency in case of proposed method. For example, if we consider two examples of clustering into 5 clusters (figures 4 and 5) we may notice, that it is possible to obtain better result (figure 4) or worse (figure 5), when one of prototypes is placed farther from the classes boundary than other ones. Clustering from figure 3 was probably also influenced by initial prototype matrix.

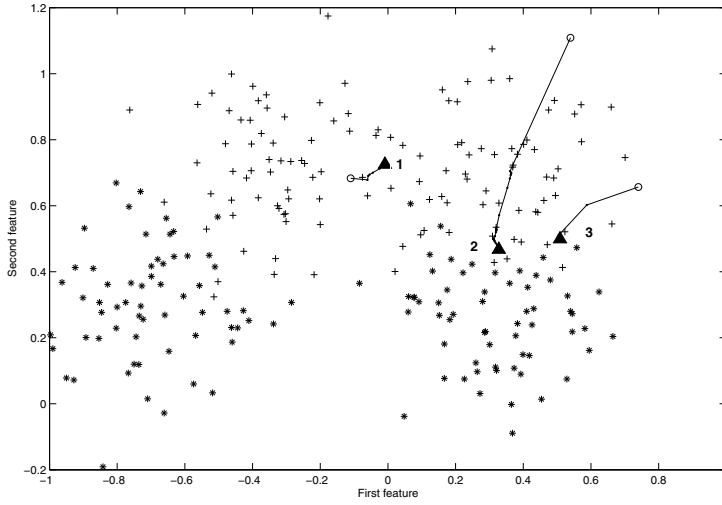


Fig. 3 Clustering the Ripley dataset into 3 clusters

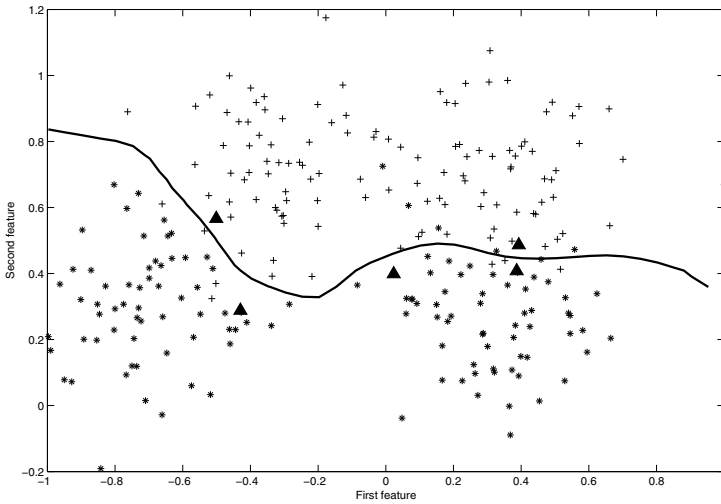


Fig. 4 Clustering the Ripley dataset into 5 clusters, influence of initial prototype matrix (better case)

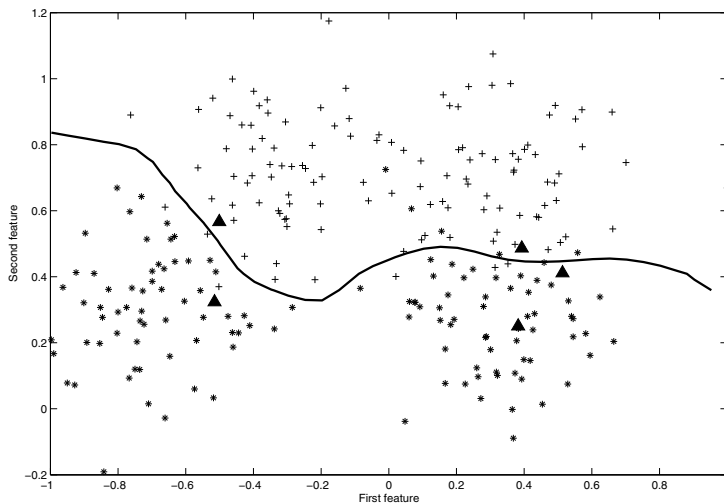


Fig. 5 Clustering the Ripley dataset into 5 clusters, influence of initial prototype matrix (worse case)

4 Conclusions

In the presented work, the new fuzzy clustering method based on minimization of the criterion function was proposed. Its goal is to find prototypes which are placed near the classes boundary. The criterion minimization procedure and its implementation were described. Presented results seem to confirm that proposed method fulfils the assumed idea - prototypes are placed near the classes boundary. Further works will be focused on designing nonlinear classifier based on prototypes obtained as a result of proposed clustering.

References

- [1] Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York (1982)
- [2] Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2, 121–167 (1998)
- [3] Duda, R.O., Hart, P.E.: Pattern classification and scene analysis. Wiley, New York (1973)
- [4] Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification. Wiley Interscience, New York (2000)
- [5] Leski, J.: Ho-Kashyap classifier with generalization control. *Pattern Recognition Letters* 24, 2281–2290 (2003)
- [6] Mangasarian, O.L., Musicant, D.R.: Lagrangian support vector machines. *Journal of Machine Learning Research* 1, 161–177 (2001)

- [7] Pedrycz, W., Waletzky, J.: Fuzzy clustering with partial supervision. *IEEE Trans. Systems, Man and Cybernetics - Part B: Cybernetics* 27, 787–795 (1997)
- [8] Ripley, B.D.: *Pattern recognition and neural networks*. Cambridge Univ. Press, Cambridge (1996)
- [9] Tou, J.T., Gonzalez, R.C.: *Pattern recognition principles*. Addison-Wesley, London (1974)
- [10] Vapnik, V.: *Statistical learning theory*. Wiley, New York (1998)

Syntactic Pattern Recognition: Survey of Frontiers and Crucial Methodological Issues

Mariusz Flasiński and Janusz Jurek

Abstract. The crucial methodological assumptions for constructing syntactic pattern recognition methods are presented in the paper. The frontiers constituting key open problems in the area of syntactic pattern recognition are identified. A survey and an analysis of the main methods based on an enhancement of context-free grammars results in formulating methodological principles for defining such methods. A discussion of key issues concerning a construction of graph grammar-based methods allows us to define methodological rules in this important area of syntactic pattern recognition.

1 Introduction

Syntactic pattern recognition [4, 11, 13, 21, 28] is one of the main approaches in the field of automatic (machine) recognition. Its main idea consists in treating a pattern as a structure of the form of string, tree or graph, and a set of structures as a formal language. A synthesis of the structures' set can be performed with the help of a formal (string, tree or graph) grammar, whereas for a recognition of an unknown structure a formal automaton is used. Complex patterns are decomposed into sub-patterns that in turn are decomposed into simpler subpatterns, etc. The elementary patterns are called *primitives* and they are represented with symbols of a formal language alphabet.

This approach has advantage over approaches based on feature vector representations, i.e. a decision-theoretic/statistical approach, neural networks, if characteristics of patterns are of a strong structural nature. Moreover, contrary to other approaches, syntactic pattern recognition methods not only perform a classification, i.e. a decision concerning a category that an unknown pattern belongs to, but also yield a structural characterization (interpretation) of this pattern. These advantages

Mariusz Flasiński · Janusz Jurek
IT Systems Department, Jagiellonian University,
ul. prof. St. Łojasiewicza 4, Cracow 30-348, Poland
e-mail: mariusz.flasinski@uj.edu.pl

were appreciated soon after constructing the first syntactic models in the sixties that resulted in rapid development of efficient methods in 1970's and 1980's in various application areas such, as: character recognition, speech recognition, scene analysis, chemical and biological structures analysis, texture analysis, fingerprint recognition, geophysics, etc. However, in the nineties a progress in a development of new methods slowed down considerably and one can say that this research area is somehow stagnating [12].

An analysis of crucial methodological issues influencing the prospects of the future development of syntactic pattern recognition is the objective of the paper. For such an analysis, firstly we should identify the main research areas bordering a domain of this theory.

Browsing through basic bibliography in the field [4, 11, 13, 21, 28], one can easily notice that, if string grammars-based methods are concerned, there is a variety of efficient solutions in case patterns can be represented with regular or context-free languages. The first problem occurs, if a set of pattern structures cannot be generated with context-free grammars (CFGs), i.e. it corresponds to a language of a context-sensitive nature, since automata for this class are of a non-polynomial time complexity. The second problem appears, if a structure of a pattern is of a "multi-dimensional" nature. Then, graph grammars are the most appropriate generative formalism. Unfortunately, a problem of parsing of graph language is, in general, NP-complete [29]. In conclusion, the following two research areas seem to constitute the frontiers of syntactic pattern recognition:

- string grammars stronger than context-free grammars generating languages with a polynomial membership problem, and
- graph grammars generating languages with a polynomial membership problem.

In the paper we will analyze both research areas in order to formulate methodological principles that should be taken into consideration while constructing related solutions. In the next section two general methodological assumptions are defined. Methodological issues concerning an enhancement of context-free grammars are discussed in the third section, whereas those that relate to constructing efficient graph grammar-based methods are presented in the fourth section. The last section contains the concluding remarks.

2 Fundamental Methodological Assumptions for Syntactic Pattern Recognition

Formulating above the main idea of syntactic pattern recognition, we have mentioned its two basic mechanisms, namely a grammar being a generator of structural representations of patterns and an automaton that is used for an analysis of these representations. In certain applications of the theory of formal languages, like e.g. a compiler design, these two formalisms are sufficient, because a grammar is designed by a human constructor who knows an idea (form) of the formal language (e.g. in case of a compiler design, it is a programming language). However, in most

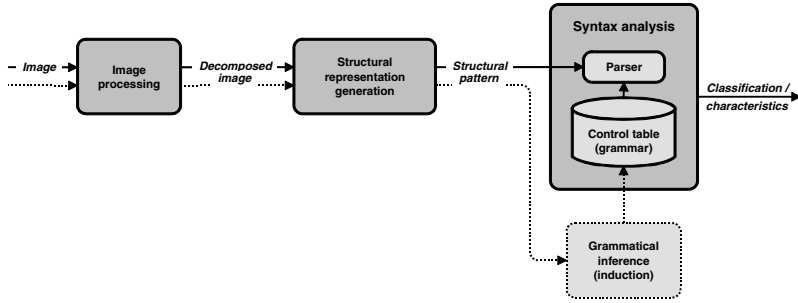


Fig. 1 A scheme of syntactic pattern recognition.

non-trivial applications in the area of pattern recognition, we do not know *explicit* a form of the language. Instead, a sample of structural patterns is given. (In fact, a sample of their representations expressed in this language is given.) In such a case an algorithm of a grammatical inference (induction) that is able to generate a grammar on the basis a sample of the language should be defined. On one hand, a problem of constructing such algorithms is even more difficult that a problem of designing automaton. On the other hand, it is claimed that a lack of such algorithms in the syntactic pattern recognition scheme (cf. Fig. 1) is one of the main obstacles for applications of syntactic methods in the pattern recognition area [12, 14]. Therefore, let us formulate the first methodological principle that allows to define a "complete" syntactic pattern recognition system shown in Fig. 1 as follows.

- I. A syntactic pattern recognition method should be complete, i.e. it should consist of: a grammar, an efficient automaton, and an efficient grammatical inference algorithm.

Classification methods of a decision-theoretic approach are based on a generic pattern representation in the form of a feature vector. Therefore, they are all-purpose in a sense they can be applied for various application areas. In contrast, constructing a syntactic method, we define a structural representation that is adequate (so, specific) for a given application area, i.e. a nature of patterns occurring in this area [4, 11, 13, 21, 28]. A form of the structural representation influences, in turn, a form (type) of a formal grammar being a basis for a construction of any method. Let us consider, now, a fundamental characterization of a formal grammar, namely its generative power. In order to define it formally, we introduce firstly, basic notions. We will make it in a general way, i.e. we do not determine a grammar structure (like a quadruple-structure for standard Chomsky's grammars), since the structure varies for grammars considered in this paper.

If Σ is a set of any symbols, then Σ^* (Kleene star) denotes a set of all strings that can be constructed by concatenating symbols of Σ , including the empty string (empty word), denoted with λ . A language L is a subset of Σ^* .

Let G be a grammar. Let components of G are denoted in the following way. V is a set of symbols (alphabet), $\Sigma \subset V$ is a set of terminal symbols, i.e. symbols that

occur in a words of a language generated with G , P is a set of productions (rules) used to generate a language, $N = V \setminus \Sigma$ is a set of nonterminal symbols, $S \in N$ is the starting symbol.

An application of a production to a string $\alpha \in V^*$ that results in obtaining a string $\beta \in V^*$ is called a derivation step, denoted $\alpha \Longrightarrow \beta$. A sequence of derivation steps (including the empty sequence) is denoted with $\xrightarrow{*}$.

The language generated with G is a set $L(G) = \{\alpha \mid S \xrightarrow{*} \alpha, \alpha \in \Sigma^*\}$.

Let X denotes a type of formal grammars. A class X of languages is a set $\mathcal{L}(X) = \{L \mid \exists G \text{ of the type } X : L = L(G)\}$, i.e. it is a set containing all the languages L that can be generated with any grammar G of the type X . We say that grammars of a type X are of a bigger generative power than grammars of a type Y , if $\mathcal{L}(Y) \subsetneq \mathcal{L}(X)$.

In general, the bigger generative power of a grammar is, the bigger computational complexity of the corresponding automaton is. Moreover, coming back to a "completeness" principle of the syntactic method that results in a necessity of constructing a grammatical inference algorithm, we may notice even bigger difficulty concerning defining such an algorithm than in case of the automaton in case of a growth of a generative power of a grammar. Summing up our considerations, we can propose the following methodological principle.

- II.** *Any syntactic pattern recognition method should be constructed for a specific problem of a strictly-defined application area, and with the use of the Ockham Razor principle with respect to generative power of an underlying grammar. That is, a grammar should be of the smallest generative power yet sufficient to generate all the possible patterns.*

3 Methodological Issues Concerning Enhancement of String Context-Free Grammars

In this section we will analyze methodological issues concerning the first syntactic pattern recognition frontier identified in an introduction, namely an enhancement of a generative power of context-free grammars.

There are two basic approaches to a construction of enhanced context-free grammars so that they are able to generate certain context-sensitive languages: *regulated rewriting (controlled grammars)* [5] developed in the theory of formal languages and *mildly context-sensitive grammars* [30] studied in the field of Natural Language Processing (NLP).

In a regulated rewriting approach a context-free grammar is devised with an additional control mechanism that allows it to control a derivation process. Indexed grammars [1] introduced by Aho in 1968 were the first class of controlled grammars. Let us define it formally.

¹ One can enhance other classes of grammars than CFGs with mechanisms discussed in the paper. Nevertheless, in syntactic pattern recognition we are interested primarily in enhancing CFGs.

Definition 1. An indexed grammar is a quintuple $G = (V, \Sigma, I, P, S)$, where: $V, \Sigma \subset V$, $S \in N$ are defined as in a previous section, I is a set of indices, P is a finite set of productions of one of the three forms:

$$(1) A \longrightarrow \alpha \text{ or } (2) A[..\] \longrightarrow B[i..\] \text{ or } (3) A[i..\] \longrightarrow [..\]\alpha ,$$

where A and $B \in N$, $i \in I$, $[..\]$ represents a stack of indices, a string in I^* , $\alpha \in V^*$.

Indices may follow any nonterminal and they are introduced in order to model a context for a derivation. A derivation in indexed grammars is defined in the following way.

1. If $A \longrightarrow X_1 \dots X_k$ is a production of type (1), then $\beta A \delta \gamma \Rightarrow \beta X_1 \delta_1 \dots X_k \delta_k \gamma$, where $\delta_i = \delta$ if $X_i \in N$ and $\delta_i = \lambda$ if $X_i \in \Sigma$.
2. If $A[..\] \longrightarrow B[x..\]$ is a production of type (2), then $\beta A \delta \gamma \Rightarrow \beta B x \delta \gamma$.
3. If $A[x..\] \longrightarrow [..\]X_1 \dots X_k$ is a production of type (3), then $\beta A x \delta \gamma \Rightarrow \beta X_1 \delta_1 \dots X_k \delta_k \gamma$, where $\delta_i = \delta$ if $X_i \in N$ and $\delta_i = \lambda$ if $X_i \in \Sigma$.

For example, let us define an indexed grammar G generating a context-sensitive language $L(G) = \{a^{2^n}, n > 0\}$. Let $G = (\{S, A, a\}, \{a\}, \{i\}, P, S)$, where P is:

1. $S[..\] \rightarrow S[i..\]$ (a production of a type 2)
2. $S \rightarrow AA$ (a production of a type 1)
3. $A[i..\] \rightarrow [..\]AA$ (a production of a type 3)
4. $A \rightarrow a$ (a production of a type 1)

Now, e.g. a string a^8 is derived in the following way.

$$\begin{aligned} S[] &\xrightarrow{1} S[i] \xrightarrow{1} S[ii] \xrightarrow{2} A[ii]A[ii] \xrightarrow{3} A[i]A[i]A[ii] \xrightarrow{3} A[i]A[i]A[i]A[i] \\ &\xrightarrow{3} A[]A[]A[i]A[i]A[i] \xrightarrow{3} \dots \xrightarrow{3} A[]A[]A[]A[]A[]A[]A[] \\ &\xrightarrow{4} aA[]A[]A[]A[]A[]A[] \xrightarrow{4} \dots \xrightarrow{4} aaaaaaaa \end{aligned}$$

As one can see, additional stacks of indices added to nonterminals are used. If a production is applied to a nonterminal with a stack, then all nonterminals of the right-hand side of the production receive copies of this stack. Such a mechanism allows us to reflect contextual dependencies during a derivation. Neither polynomial parser nor polynomial grammatical inference algorithm have been defined for indexed grammars.

Let us notice that some additional syntactic constructs (brackets: $[\]$, indices) occur in both grammar productions and non-final phrases of a derivation, apart from terminal and nonterminal symbols. These constructs do not occur in words of a language generated and they play a role of *operators* controlling a derivation. An occurrence of such operators is typical for mildly context-sensitive grammars (MCSGs) [30] used in NLP and mentioned above. Mildly context-sensitive languages (MCSLs) fulfill the following properties. MCSLs contain context-free languages and certain languages with context dependencies ($L_1 = \{a^n b^n c^n \mid n \geq 0\}$, $L_2 = \{a^n b^m c^n d^m \mid n, m \geq 0\}$, $L_3 = \{ww \mid w \in \{a, b\}^*\}$). Their membership problem is solvable in a deterministic polynomial time. MCSLs have the linear growth property (if strings of a language are ordered in a sequence according to their length, then two successive lengths do not differ in arbitrary large amounts). The best known

MCSGs include: *linear indexed grammars* being a simplified version of indexed grammars, *head grammars* allowing one to combine derived substrings with the help of their special markers called *heads*, and *combinatory categorial grammars* that use rules containing also operators, which do not occur in words of a language². Polynomial parsing algorithms for MCSGs have been defined. However, although these formalisms have been introduced in mid 1980's, grammatical inference algorithms have not been constructed for them till now.

Let us come back to a regulated rewriting approach. The second, apart from indexing, popular derivation control method consists in adding a separate control mechanism to a grammar production (i.e. the mechanism is not "hidden" in left- or right-hand sides of a production). Such a methodology is used in programmed grammars [24] introduced by Rosenkrantz in 1969.

Definition 2. A programmed grammar is a quintuple $G = (V, \Sigma, J, P, S)$, where: $V, \Sigma \subset V, S \in N$ are defined as in a previous section, J is a set of production labels, P is a finite set of productions of the form :

$$(r) \alpha \longrightarrow \beta \ S(U) \ F(W), \text{ in which}$$

$\alpha \longrightarrow \beta, \alpha \in V^*NV^*, \beta \in V^*$, is called the core, (r) is the production label, $r \in J, U \subset J$ is the success field and $W \subset J$ is the failure field.

A derivation is defined as follows. A production labeled with (1) is applied firstly. If it is possible to apply a production (r) , then after its application the next production is chosen from its success field U . Otherwise, we choose the next production from the failure field W .

Let us define a programmed grammar G such that $L(G) = \{a^n b^n c^n, n > 0\}$. Let $G = (\{S, A, B, C, a, b, c\}, \{a, b, c\}, \{1, \dots, 7\}, P, S)$, where P is:

- | | |
|---|--|
| 1. $S \rightarrow ABC \ S(\{2, 5\}) \ F(\emptyset)$ | 5. $A \rightarrow a \ S(\{6\}) \ F(\emptyset)$ |
| 2. $A \rightarrow aA \ S(\{3\}) \ F(\emptyset)$ | 6. $B \rightarrow b \ S(\{7\}) \ F(\emptyset)$ |
| 3. $B \rightarrow bB \ S(\{4\}) \ F(\emptyset)$ | 7. $C \rightarrow c \ S(\{\emptyset\}) \ F(\emptyset)$ |
| 4. $C \rightarrow cC \ S(\{2, 5\}) \ F(\emptyset)$ | |

For example, a derivation of a string $aabbcc$ is performed as follows.

$$\begin{aligned} S &\xrightarrow{1} ABC \xrightarrow{2} aABC \xrightarrow{3} aAbBC \xrightarrow{4} aAbBcC \\ &\xrightarrow{5} aabBcC \xrightarrow{6} aabbcC \xrightarrow{7} aabbcc \end{aligned}$$

A static control mechanism of programmed grammars (success and failure fields possess fixed indices of productions) has been extended in DPLL(k) grammars (Dynamically Programmed LL(k) grammars) [9, 16, 17]. Instead of success and failure fields, every production is devised with a control tape. A head of a tape can write indices of productions, read and move. A derivation is performed according to a content of a tape. Additionally, a derivation is deterministic in a sense of standard LL(k) grammars [25] and "recursive steps" are limited. Parsing algorithm is of $O(n)$

² *Tree adjoining grammars* are the fourth well-known MCSGs, but, in fact, they are tree (not string) grammars.

computational complexity, and a grammatical inference algorithm is also of polynomial complexity, $O(m^3 \cdot n^3)$, where m is a sample size, n is the maximum length of a string in a sample.

Apart from the inference algorithm mentioned above, there is few inference methods that can handle some subset of context-sensitive languages. In a unique syntactic pattern recognition model presented in [2], a pattern language is represented with *augmented regular expressions* and a formal grammar is not used at all. An efficient parsing algorithm and a learning method have been defined in this method. In the area of NLP, a polynomial-time algorithm of grammatical inference has been defined recently for Simple External Contextual ($\mathcal{S}\mathcal{E}\mathcal{C}_p$) grammars [19] belonging to the Marcus (non-Chomskyan) paradigm of contextual grammars [18]. In both cases, defining a language with contextual dependencies does not involve an enhancement of productions of the Chomskyan context-free form.

A short survey of the basic models of enhancement of context-free grammars shows that a definition of a grammatical inference algorithm is the main problem in this area. In most enhanced models, such as indexed grammars, head grammars and combinatory grammars, a derivation control mechanism is "hidden" (cleverly) in grammar productions. It is made with the use of some syntax constructs like stacks, heads, operators that do not occur in the words of a language. Let us notice that the main idea of standard inference methods (i.e. used for regular and context-free languages) consists in looking for similarities among sample strings. An alphabet of a grammar inferred contains only terminal symbols that occur in the language sample and nonterminals that are entered in a (relatively) simple way as "classes of abstractions" for certain substrings. The only "operator" used is a simple *catenation operator*. In a way, this operator is "visible" in a derived word. However, if grammar productions contain operators that disappear during a derivation and do not occur in a derived word, a grammatical inference problem becomes very difficult, since a syntax of sample words does not deliver any information related to a history of obtaining these words with such operators. It is hardly likely that algorithms reconstructing such a history can be of a polynomial complexity. Thus, we can formulate the following methodological principle.

III. *Defining a control mechanism enhancing a grammar as a separate element makes a development of an efficient grammatical inference algorithm easier than "hiding" this mechanism in left- or right-hand sides of productions with the help of additional syntactic operators.*

4 Methodological Issues Concerning Graph Grammars

Graph grammars are widely used for a synthesis (generation) of multi-dimensional pattern structures. A possibility of their use in machine vision and syntactic pattern recognition depends on a balance between a generative power sufficient for representing patterns and a parsing efficiency. Therefore, although the first graph automata were proposed in the late 1960's, only a few graph grammar-based syntactic pattern recognition models have been presented for last forty years. Web

automata were proposed by Rosenfeld and Milgram in 1972 [23]. An efficient parser for expansive graph grammars was constructed by Fu and Shi in 1983 [26]. In 1990 two parsing algorithms for plex grammars were defined independently by Bunke and Haller [3], and Peng, Yamamoto and Aoki [22]. In the early 1990's two parsing methods for relational grammars were proposed [6, 31]. An efficient, $O(n^2)$, parser for the ETPL(k) subclass of the well-known edNLC [15] graph grammars was constructed in 1993 [7, 8]. A parsing method for reserved graph grammars was proposed in 2001 [32].

An analysis of the bibliography mentioned above shows that the methods have been developed for precisely defined (sub)classes of graph grammars and their application area has been determined by a nature of graph structures belonging to a specific graph language. As we have mentioned it in introduction, a membership problem for graph languages is, in general, NP-complete [29]. Therefore, in case of a construction of a graph parser, either a generative power of a class of grammars is usually decreased with imposing certain restrictions or one defines a specific class of graph languages with a polynomial membership problem (like e.g. in case of expansive grammars [26]). Summing up, we can define the following principle.

IV. *During a process of defining a class of graph grammar one should focus primarily on polynomial decidability of a membership problem for a language generated.*

In case of a graph grammar inference area in a syntactic pattern recognition, the efficient method has been defined only for the parsable ETPL(k) grammars [7, 8] mentioned above [10]. It results from the fact that this problem is more complex and difficult than a parsing problem. Analyzing the inference algorithm for ETPL(k) grammars, one can easily see that it takes advantage of two analogies of string grammars. Firstly, deterministic properties of a derivation of ETPL(k) graph languages are obtained on the analogy of the well-known string deterministic LL(k) grammars [25]. In fact, it allows us not only to defined an efficient grammatical inference algorithm, but also an efficient parser. Secondly, a general scheme of an inference of an ETPL(k) grammar is analogous to a scheme of formal derivatives technique used for string grammars [11]. Looking for analogies in the area of string languages seems to be a good methodological technique while we make a research in the area of graph languages. Thus, let us formulate the following principle.

V. *During a process of constructing a graph automaton and an algorithm of a graph grammar inference one should look for analogical constructs and mechanisms in an area of the theory of string languages.*

5 Conclusions

Theory of formal languages (mathematical linguistics) constitutes a formal basis for various research areas in computer science. A specificity of an area determines methodological principles that should be followed during a research. In other words, for various areas various principles are valid. The general principles concerning an

application of mathematical linguistics formalisms in a syntactic pattern recognition area have been formulated in the paper. Let us remember that they are not necessarily valid in other research areas that make use of formalisms of mathematical linguistics such, as e.g. natural language processing, compiler design.

Syntactic pattern recognition paradigm is primarily an approach in a machine recognition area. However, it relates also strongly with the Artificial Intelligence (AI) field, since it can be used not only for pattern classification, but for pattern understanding, as well (see e.g. [27]). Therefore, a set of methodological principles presented in the paper will be extended in the future with new ones connected with the AI paradigms.

References

- [1] Aho, A.V.: Indexed grammars - an extension of context-free grammars. *Journal of the Association for Computing Machinery* 15, 647–671 (1968)
- [2] Alquezar, R., Sanfeliu, A.: Recognition and learning of a class of context-sensitive languages described by augmented regular expressions. *Pattern Recognition* 30, 163–182 (1997)
- [3] Bunke, H.O., Haller, B.: A parser for context free plex grammars. In: Nagl, M. (ed.) *WG 1989. LNCS*, vol. 411, pp. 136–150. Springer, Heidelberg (1990)
- [4] Bunke, H.O., Sanfeliu, A. (eds.): *Syntactic and Structural Pattern Recognition - Theory and Applications*. World Scientific, Singapore (1990)
- [5] Dassow, J., Paun, G.: *Regulated Rewriting in Formal Language Theory*. Springer, New York (1990)
- [6] Ferruci, F., Tortora, G., Tucci, M., Vitiello, G.: A predictive parser for visual languages specified by relational grammars. In: *Proc. IEEE Symp. Visual Lang. VL 1994*, pp. 245–252 (1994)
- [7] Flasiński, M.: On the parsing of deterministic graph languages for syntactic pattern recognition. *Pattern Recognition* 26, 1–16 (1993)
- [8] Flasiński, M.: Power properties of NLC graph grammars with a polynomial membership problem. *Theoretical Computer Science* 201, 189–231 (1998)
- [9] Flasiński, M., Jurek, J.: Dynamically programmed automata for quasi context sensitive languages as a tool for inference support in pattern recognition-based real-time control expert systems. *Pattern Recognition* 32, 671–690 (1999)
- [10] Flasiński, M.: Inference of parsable graph grammars for syntactic pattern recognition. *Fundamenta Informaticae* 80, 379–413 (2007)
- [11] Fu, K.S.: *Syntactic Pattern Recognition and Applications*. Prentice Hall, Englewood Cliffs (1982)
- [12] Goldfarb, L.: Pattern representation and the future of pattern recognition. In: *Proc. ICPR 2004 Workshop*, Cambridge UK, August 22 (2004)
- [13] Gonzales, R.C., Thomason, M.G.: *Syntactic Pattern Recognition: An Introduction*. Addison-Wesley, Reading (1978)
- [14] Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: a review. *IEEE Trans. Pattern Anal. Machine Intell.* 22, 4–37 (2000)
- [15] Janssens, D., Rozenberg, G.: On the structure of node-label-controlled graph languages. *Information Sciences* 20, 191–216 (1980)

- [16] Jurek, J.: On the linear computational complexity of the parser for quasi context sensitive languages. *Pattern Recognition Letters* 21, 179–187 (2000)
- [17] Jurek, J.: Recent developments of the syntactic pattern recognition model based on quasi-context sensitive languages. *Pattern Recognition Letters* 26, 1011–1018 (2005)
- [18] Marcus, S.: Contextual grammars. *Rev. Roum. Math. Pures Appl.* 14, 1525–1534 (1969)
- [19] Oates, T., Armstrong, T., Becerra-Bonache, L., Atamas, M.: Inferring grammars for mildly context sensitive languages in polynomial time. In: Sakakibara, Y., Kobayashi, S., Sato, K., Nishino, T., Tomita, E. (eds.) *ICGI 2006. LNCS (LNAI)*, vol. 4201, pp. 137–147. Springer, Heidelberg (2006)
- [20] Okhotin, A.: Boolean grammars. *Information and Computation* 194, 19–48 (2004)
- [21] Pavlidis, T.: *Structural Pattern Recognition*. Springer, New York (1977)
- [22] Peng, K.J., Yamamoto, T., Aoki, Y.: A new parsing scheme for plex grammars. *Pattern Recognition* 23, 393–402 (1990)
- [23] Rosenfeld, A., Milgram, D.L.: Web automata and web grammars. *Machine Intelligence* 7, 307–324 (1972)
- [24] Rosenkrantz, D.J.: Programmed grammars and classes of formal languages. *Journal of the Association for Computing Machinery* 16, 107–131 (1969)
- [25] Rosenkrantz, D.J., Stearns, R.E.: Properties of deterministic top-down grammars. *Information Control* 17, 226–256 (1970)
- [26] Shi, Q.Y., Fu, K.S.: Parsing and translation of attributed expansive graph languages for scene analysis. *IEEE Trans. Pattern Analysis Mach. Intell.* 5, 472–485 (1983)
- [27] Tadeusiewicz, R., Ogiela, M.R.: *Medical Image Understanding Technology*. Springer, Heidelberg (2004)
- [28] Tanaka, E.: Theoretical aspects of syntactic pattern recognition. *Pattern Recognition* 28, 1053–1061 (1995)
- [29] Turan, G.: On the complexity of graph grammars. Rep. Automata Theory Research Group. Szeged (1982)
- [30] Vijay-Shanker, K., Weir, D.J.: The equivalence of four extensions of context-free grammars. *Mathematical Systems Theory* 27, 511–546 (1994)
- [31] Wittenburg, K.: Earley-style parsing for relational grammars. In: *Proc. IEEE Symp. Visual Lang. VL 1992*, pp. 192–199 (1992)
- [32] Zhang, D.Q., Zhang, K., Cao, J.: A context-sensitive graph grammar formalism for the specification of visual languages. *The Computer Journal* 44, 186–200 (2001)

New Measures of Classifier Competence – Heuristics and Application to the Design of Multiple Classifier Systems

Bartłomiej Antosik and Marek Kurzynski

Abstract. In the paper three new methods based on different heuristics for calculating competence of a classifier are proposed. In the common two-step procedure, first the so-called source competence at validation points are determined and next these competence values are extended to the entire feature space. The first proposition of the source competence reflects both the uncertainty of classifier's decision and its correctness. In the second method the source competence states the difference of membership degrees to the fuzzy sets of competent and incompetent classifiers. The third method is based on the normalized entropy of supports which classifier gives for particular classes. The dynamic selection (DCS) and dynamic ensemble selection (DES) systems were developed using proposed measures of competence. The performance of multiclassifiers was evaluated using six benchmark databases from the UCI Machine Learning Repository. Classification results obtained for five multiclassifier system with selection and fusion strategy were used for a comparison. The experimental results showed that, regardless of the strategy used by the multiclassifier system, the classification accuracy for homogeneous base classifiers has increased when the measure of competence was employed.

1 Introduction

Multiple classifier systems (MCS) which combines responses of set of classifiers can achieved better classification accuracy than any single classifier in the ensemble. This is because different classifiers may offer complementary information about the objects to be classified. For the classifiers combination two main approaches used are classifiers fusion and classifiers selection [10]. In the first method, all classifiers in the ensemble contribute to the decision of MCS. In the second approach, a single

Bartłomiej Antosik · Marek Kurzynski

Technical University of Wrocław,

Faculty of Electronics, Chair of Systems and Computer Networks, Wyb. Wyspińskiego 27,
50-370 Wrocław, Poland

e-mail: [bartlomiej.antosik,marek.kurzynski}@pwr.wroc.pl](mailto:{bartlomiej.antosik,marek.kurzynski}@pwr.wroc.pl)

classifier is selected from the ensemble and its decision is treated as the decision of the MCS. The selection of a classifier can be either static or dynamic. In the first method selection is a static process carried out only once for all test patterns, in the second approach however, selection is a dynamic process repeated for different test patterns [9].

Recently, dynamic ensemble selection methods (DES) has been developed which first dynamically select an ensemble of classifiers from the entire set (pool) and then combine the selected classifiers by majority voting. The most dynamic classifier (ensemble of classifiers) selection schemes use the concept of classifier competence on a defined neighbourhood or region [10], such as the local accuracy estimation [3, 12, 16], Bayes confidence measure [8], multiple classifier behaviour [7] or randomized classification model [15], among others.

In the approach presented in [13, 14, 15], the competence measure of classifier is determined via the two-step procedure, in which first the so-called source competence at validation points are determined and next these competence values are extended to the entire feature space using potential function model.

In this paper we present new concepts of measure of source competence based on the different heuristics. In the first method the source competence is calculated as a common measure of uncertainty of classifier's decision and its correctness. The second method is based on the fuzzy logic and the source competence states the difference between membership degrees of classifier to the fuzzy sets of competent and incompetent classifiers. The third method is based on the entropy of supports which classifier gives for particular classes.

This paper is divided into five sections and organized as follows. In Section 2 three different methods for determining measure of classifier competence are presented. In Section 3 two multiclassifier systems using proposed measures of classifier competence are developed. Computer experiments are described in Section 4 and Section 5 concludes the paper.

2 Theoretical Framework

2.1 Preliminaries

Consider an n -dimensional feature space $\mathcal{X} \subseteq \mathcal{R}^n$ and a set of class numbers $\mathcal{M} = \{1, 2, \dots, M\}$. Let

$$\psi: \mathcal{X} \rightarrow \mathcal{M} \quad (1)$$

be a trained classifier which produces a set of discriminant functions $(d_1(x), d_2(x), \dots, d_M(x))$ for a given object described by a feature vector x . The value of the discriminant function $d_i(x)$, $i = 1, 2, \dots, M$ represents a support given by the classifier ψ for the i -th class. Without loss of generality we assume that $d_i(x) > 0$ and $\sum d_i(x) = 1$. Classification is made according to the maximum rule, i.e.

$$\psi(x) = i \Leftrightarrow d_i(x) = \max_{k \in \mathcal{M}} d_k(x). \quad (2)$$

We assume that a validation dataset is available:

$$V_N = \{(x_1, j_1), (x_2, j_2), \dots, (x_N, j_N)\}, \quad (3)$$

where $x_k \in \mathcal{X}$, $k = 1, 2, \dots, N$ denotes the feature vector representing the k -th object in the dataset and $j_k \in \mathcal{M}$ denotes the object's class label.

Now, using the set (3) we determine a competence measure $C(\psi|x)$ which evaluates the competence of classifier ψ , i.e. its capability to correct activity (correct classification) at a point $x \in \mathcal{X}$.

2.2 The Measures of Classifier Competence

According to the method presented in [13, 14], the competence $C(\psi|x)$ of the classifier ψ for the object x is calculated in the following two-step procedure. First, the so-called source competence at validation points $C_{SRC}(\psi|x_k)$, ($x_k \in V$) are determined and next these competence values are extended to the entire feature space \mathcal{X} using normalized Gaussian potential function model [10]:

$$C(\psi|x) = \frac{\sum_{x_k \in V} C_{SRC}(\psi|x_k) \exp(-\text{dist}(x, x_k)^2)}{\sum_{x_k \in V} \exp(-\text{dist}(x, x_k)^2)}, \quad (4)$$

where $\text{dist}(x, x_k)$ is the Euclidean distance between the objects x and x_k .

In the next subsections three different methods for the calculation of the source competence $C_{SRC}(\psi|x_k)$ are proposed. The methods developed satisfy the following properties formulated in [13]:

1. $-1 \leq C_{SRC}(\psi|x_k) \leq 1$,
2. $C_{SRC}(\psi|x_k)$ is a strictly increasing function of $d_{j_k}(x_k)$,
3. if $C_{SRC}(\psi|x_k) < 0$ then ψ is incompetent,
4. if $C_{SRC}(\psi|x_k) > 0$ then ψ is competent,
5. if $C_{SRC}(\psi|x_k) = 0$ then ψ is neutral.

2.2.1 The Minimal Difference (MD) Method

In the classification procedure (1) values of discriminant functions, that is supports given by the classifier ψ for particular classes, describe the confidence of the classifier's decision. Extreme case when one support is equal to 1 and remaining supports are equal to 0 denotes the absolutely firm classifier's decision. In typical cases however, when values of all discriminant functions are between 0 and 1, we deal with an unreliable decision.

The MD method taking into account the uncertainty of classifier's decision as well as its correctness, leads to the following formula for the source competence of classifier ψ at a validation point $x_k \in V$:

$$C_{SRC}(\psi|x_k) = \min_{j \in \mathcal{M}, j \neq j_k} [d_{j_k}(x_k) - d_j(x_k)]. \quad (5)$$

If classifier ψ correctly classifies a validation point x_k , then $C_{SRC}(\psi|x_k) > 0$ and ψ is considered as competent. If classifier ψ makes an error, then $C_{SRC}(\psi|x_k) < 0$ and ψ is incompetent. In both cases the value of competence (incompetence) depends on the uncertainty of the classifier's decision.

2.2.2 The Fuzzy Competence (FC) Method

In the terms of two-valued logic the set of classifiers correctly classifying a validation point x_k ($\psi(x_k) = j_k$) can be considered as a crisp set of competent classifiers $\mathcal{C}(x_k)$ at a point $x_k \in V$. Similarly, classifiers which incorrectly classify a point x_k ($\psi(x_k) = j$ ($j \neq j_k$)) create the set of incompetent classifiers $\mathcal{I}\mathcal{C}(x_k)$ at a point $x_k \in V$. It means that both sets are disjunctive and any classifier or belongs to $\mathcal{C}(x_k)$ (does not belong to $\mathcal{I}\mathcal{C}(x_k)$) or does not belong to $\mathcal{C}(x_k)$ (belongs to $\mathcal{I}\mathcal{C}(x_k)$). In other words, for competent or incompetent classifier characteristic functions of both sets are equal $\lambda_{\mathcal{C}}(\psi) = 1$ and $\lambda_{\mathcal{I}\mathcal{C}}(\psi) = 0$ or $\lambda_{\mathcal{I}\mathcal{C}}(\psi) = 1$ and $\lambda_{\mathcal{C}}(\psi) = 0$, respectively.

This concept can be easily generalized and presented in the terms of fuzzy logic. Now, $\mathcal{C}(x_k)$ as well as $\mathcal{I}\mathcal{C}(x_k)$ are fuzzy sets with the following membership functions:

$$\mu_{\mathcal{C}(x_k)}(\psi) = \frac{d_{j_k}(x_k) - \min_{j \in \mathcal{M}} d_j(x_k)}{\max_{j \in \mathcal{M}} d_j(x_k) - \min_{j \in \mathcal{M}} d_j(x_k)}, \quad (6)$$

$$\mu_{\mathcal{I}\mathcal{C}(x_k)}(\psi) = \frac{\max_{j \in \mathcal{M}} d_j(x_k) - d_{j_k}(x_k)}{\max_{j \in \mathcal{M}} d_j(x_k) - \min_{j \in \mathcal{M}} d_j(x_k)}, \quad (7)$$

where $d_j(x_k)$ ($j \in \mathcal{M}$) are supports (values of discriminant functions) produced by the classifier ψ at a validation point x_k ($d_{j_k}(x_k)$ is the support for correct class). It is worth noting, that both sets are defined in the real-valued space and the membership degree of the classifier ψ into $\mathcal{C}(x_k)$ and $\mathcal{I}\mathcal{C}(x_k)$ depends on the support given by ψ for the correct class.

In the FC method, the source competence of a classifier ψ at a point $x_k \in V$ is defined as the following difference:

$$C_{SRC}(\psi|x_k) = \mu_{\mathcal{C}(x_k)}(\psi) - \mu_{\mathcal{I}\mathcal{C}(x_k)}(\psi). \quad (8)$$

From (6) and (7) after simple calculations we get

$$C_{SRC}(\psi|x_k) = \frac{2 \cdot d_{j_k}(x_k) - \max_{j \in \mathcal{M}} d_j(x_k) - \min_{j \in \mathcal{M}} d_j(x_k)}{\max_{j \in \mathcal{M}} d_j(x_k) - \min_{j \in \mathcal{M}} d_j(x_k)}. \quad (9)$$

In Fig. 1 membership functions (6), (7) and the source competence (9) of the classifier ψ at a point x_k are plotted against support given by ψ for the correct class.

2.2.3 The Entropy (ENT) Method

In this method the source competence of classifier ψ at the validation point x_k is a product of two factors, which separately determine the sign and absolute value

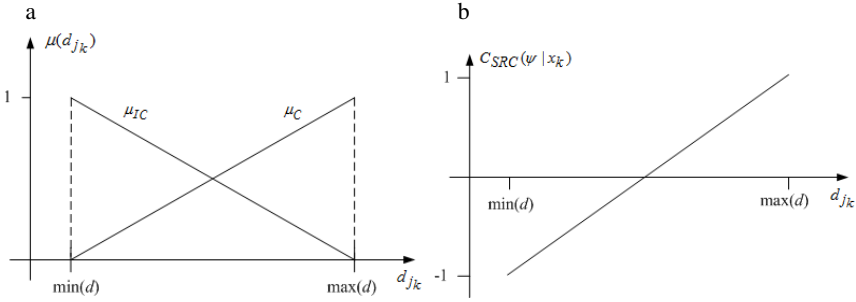


Fig. 1 The membership functions of sets $\mathcal{C}(x_k)$ and $\mathcal{I}\mathcal{C}(x_k)$ (a) and the source competence of ψ (b) plotted against the support for the correct class.

of the competence. The value of the source competence is inverse proportional to the normalized entropy of its supports vector and the sign of competence is simply determined by correct/incorrect classification of x_k . This idea leads to the following proposition of the ENT method:

$$C_{SRC}(\psi|x_k) = (-1)^{I_{\{\psi(x_k) \neq i_k\}}} \cdot [1 - \frac{1}{\log n} \sum_{i=1}^M d_i(x_k) \log d_i(x_k)] \quad (10)$$

where $I_{\{A\}}$ is the indicator of the set (event) A .

3 Dynamic Classifier Selection Systems

The measures of competence can be incorporated in any multiclassifier system providing that \mathcal{X} is an metric space.

Let us assume that a set (pool) of trained base classifiers $\mathcal{L} = \{\psi_1, \psi_2, \dots, \psi_L\}$ and the validation dataset V_N are given.

We define two multiclassifier systems based on proposed measures of competence, each one employing different strategy.

1. The multiclassifier $F_1(x)$ is the classifier with the highest positive competence value at the point x :

$$F_1(x) = \psi_l(x) \Leftrightarrow C(\psi_l|x) > 0 \wedge C(\psi_l|x) = \max_{k=1,2,\dots,L} C(\psi_k|x). \quad (11)$$

The multiclassifier $F_1(x)$ uses a selection strategy, i.e. for each object described by a feature vector x it selects a single classifier to be used for classification.

2. The multiclassifier $F_2(x)$ represents an ensemble selection approach where first the set of competent classifiers is found and next the selected classifiers are combined by weighted majority voting where the weights are equal to the competence values. Thus, we have:

$$F_2(x) = j \Leftrightarrow d_j^{(F_2)}(x) = \max_{k \in \mathcal{M}} d_k^{(F_2)}(x), \quad (12)$$

where

$$d_j^{(F_2)}(x) = \sum_{\psi_l: C(\psi_l|x) > 0} C(\psi_l|x) d_{jl}(x). \quad (13)$$

4 Experiments

4.1 Benchmark Data and Experimental Setup

Benchmark databases used in the experiments were obtained from the UCI Machine Learning Repository [2]. A brief description of each database is given in Table 1.

Table 1 A brief description of the data sets used

Dataset	# objects	# features	# classes
Iris	150	4	3
Wine	178	13	3
Breast	699	9	2
Glass	214	9	6
Ionosphere	351	34	2
Vowel	990	10	11

Selected databases represent classification problems with objects described by continuous feature vectors. For each database, feature vectors were normalized for zero mean and unit standard deviation. The training and testing datasets were extracted from each database using two-fold cross-validation. The training set was also used as the validation set.

The experiments were conducted in MATLAB using PRTools [4].

4.2 Base Classifiers

Two types of classifier ensembles were used in the experiments: homogeneous and heterogeneous. The homogeneous ensemble consisted of 50 unpruned decision tree classifiers with misclassification (purity) splitting criterion. Each classifier was trained using bootstrapping of the initial training set.

The pool of heterogeneous base classifiers using in the experiments, consisted of the following nine classifiers [5]:

1. LDC — linear discriminant classifier based on normal distributions with the same covariance matrix for each class
2. QDC — quadratic discriminant classifier based on normal distributions with different covariance matrix for each class
3. NMC — nearest mean classifier
4. k-NN — k -nearest neighbours classifier with $k = 1, 5, 10$

Table 2 Classification accuracies of the MCSs using heterogeneous ensembles. The best result for each database is bolded. The statistically significant differences found are given under the classification accuracies.

Database / Mean accuracy [%]							
No	MCS	Iris	Wine	Breast	Glass	Iono	Vowel
1	SB	95.20	97.86	95.62	69.06	85.58	94.75
2	MV	94.93	96.40	96.37	66.37	87.18	87.66
3	DCS	94.27	97.64	95.71	66.54	86.15	90.65
4	DCS-LA	96.53	97.41	95.91	65.91	86.21	85.51
5	DES-MCB	95.60	95.51	96.38	69.54	85.81	75.78
6	F_1 -MD	92.13	92.37	95.39	61.50	78.92	77.86
7	F_1 -FC	94.93	96.50	96.02	66.35	83.47	84.22
8	F_1 -ENT	92.67	92.68	95.19	57.44	78.12	79.76
9	F_2 -MD	95.33	95.84	95.48	62.24	85.47	78.79
10	F_2 -FC	95.47	96.40	95.48	64.20	91.51	90.67
11	F_2 -ENT	95.07	95.83	96.17	53.64	86.50	88.24

- 5. PARZEN — Parzen classifier with the Gaussian kernel and optimal smoothing parameter h_{opt} and smoothing parameter $h_{opt}/2$
- 6. TREE — Unpruned tree classifier with purity splitting criterion

4.3 Multiple Classifier Systems

The F_1 and F_2 systems with MD, FC and ENT competence methods were compared against five multiclassifier systems:

- 1. SB system – This system selects the single best classifier in the pool;
- 2. MV system – This system is based on majority voting of all classifiers in the pool;
- 3. DCS system – This system defines competence of the classifier ψ at the test object x using potential function model [11] with source competence at validation point equal to +1 if this point is correctly classified or -1 if error is made.
- 4. DCS-LA system – This system defines the competence of the classifier ψ for the test object x as the local classification accuracy. The accuracy is estimated using the k -nearest neighbours of the test sample x taken from the validation set V . $k = 10$ was chosen since for this value the DCS-LA system had the best overall performance in previous studies [16].
- 5. DES-MCB system – This system is the same as the DCS system, except the competence is defined using the concept of multiple classifier behaviour [7].

4.4 Results and Discussion

Classification accuracies (i.e. the percentage of correctly classified objects) were calculated for the MCSs as average values obtained over 10 runs (5 replications of two-fold cross validation). The results obtained for the MCSs using heterogeneous and homogeneous ensembles are shown in Table 2 and Table 3, respectively. Statistical differences between the performances of the F_1 and F_2 systems and the five MCSs were evaluated using 5x2cv F test [11] and results are given under the classification accuracies.

Table 3 Classification accuracies of the MCSs using homogeneous ensembles. The best result for each database is bolded. The statistically significant differences found are given under the classification accuracies.

		Database / Mean accuracy [%]					
No	MCS	Iris	Wine	Breast	Glass	Iono	Vowel
1	SB	91.07	86.77	94.68	65.41	86.38	60.32
2	MV	92.53	93.17	96.08	74.12	89.46	82.46
3	DCS	91.20	86.56	94.59	65.74	87.12	64.87
4	DCS-LA	90.80	85.31	94.45	65.33	84.44	62.48
5	DES-MCB	92.93	93.17	96.11	74.03	89.17	83.25
6	F_1 -MD	92.80	86.89	95.39	65.53	86.49	73.09
		1,3,4	4			4	1,3,4
7	F_1 -FC	90.13	84.50	94.31	64.02	85.30	61.90
8	F_1 -ENT	92.27	86.66	95.54	64.51	86.55	71.25
		1,3,4	4	1,4		4	1,3,4
9	F_2 -MD	93.60	93.06	96.02	73.38	92.65	86.85
		1,2,3,4	1,3,4	1,3,4	1,3,4	1,2,3,4,5	1,2,3,4,5
10	F_2 -FC	93.20	93.07	95.99	73.75	92.48	86.00
		1,3,4	1,3,4	1,3,4	1,3,4	1,2,3,4,5	1,2,3,4,5
11	F_2 -ENT	93.60	93.62	96.05	73.37	92.71	87.31
		1,2,3,4	1,3,4	1,3,4	1,3,4	1,2,3,4,5	1,2,3,4,5

These results imply the following conclusions:

1. The best F_1 system outperformed the SB classifier by 1.73%, 0.12%, 0.86%, 0.12%, 0.17% and 12.77% for homogeneous ensembles and for different databases, respectively;
2. The best F_2 system outperformed the SB classifier by 2.53%, 6.85%, 1.37%, 7.97%, 6.43% and 16.99% for homogeneous ensembles and for different databases, respectively;
3. The F_2 systems achieved the highest classification accuracy for 1 dataset and 4 datasets for heterogeneous and homogeneous ensembles, respectively;
4. The F_1 and F_2 systems produced statistically significant higher scores in 34 out of 60 cases and 11 out of 60 cases for homogeneous and heterogeneous classifiers, respectively;

5. The F_1 systems produced statistically significant higher scores against DCS systems in 4 out of 12 cases for homogeneous classifiers;
6. The F_2 systems produced statistically significant higher scores against DES system in 2 out of 6 cases for homogeneous and heterogeneous classifiers.

5 Conclusions

In the paper new methods for calculating the competence of a classifier in the feature space were presented. Two multiclassifier systems incorporating the competence were evaluated using six databases from the UCI Machine Learning Repository.

The reason for the better performance of the F_1 and F_2 systems for the homogeneous ensemble could be that classifiers in the ensemble were diverse and their classification accuracies varied noticeably within the ensemble. Therefore, a wide range of competences were obtained and the classifiers evaluated as better-than-random could be successfully selected by the system developed.

The relatively worse performance of the F_1 and F_2 systems for the heterogeneous ensemble could be explained by the fact that most of classifiers in the ensemble were highly accurate. For example, the SB system outperformed at least one of the other 10 MCSs for all 6 databases.

References

- [1] Alpaydin, E.: Combined 5 x 2 cv F test for comparing supervised classification learning algorithms. *Neural Computation* 11, 1885–1892 (1999)
- [2] Asuncion, A., Newman, D.: UCI Machine Learning Repository. Department of Information and Computer Science. University of California, Irvine (2007), <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [3] Didaci, L., Giacinto, G., Roli, F., Marcialis, G.: A study on the performances of dynamic classifier selection based on local accuracy estimation. *Pattern Recognition* 38, 2188–2191 (2005)
- [4] Duin, R., Juszczak, P., Paclik, P., Pekalska, E., Ridder, D., Tax, D.: PR-Tools 4.1, A Matlab Toolbox for Pattern Recognition, Delft University of Technology (2007), <http://prtools.org>
- [5] Duda, R., Hart, P., Stork, G.: *Pattern Classification*. John Wiley and Sons, New York (2000)
- [6] Eulanda, M., Santos, D., Sabourin, R., Maupin, P.: A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. *Pattern Recognition* 41, 2993–3009 (2008)
- [7] Giacinto, G., Roli, F.: Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recognition* 34, 1879–1881 (2001)
- [8] Huenupan, F., Yoma, N., et al.: Confidence based multiple classifier fusion in speaker verification. *Pattern Recognition Letters* 29, 957–966 (2008)
- [9] Ko, A., Sabourin, R., Britto, A.: From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition* 41, 1718–1733 (2008)
- [10] Kuncheva, I.: *Combining Pattern Classifiers: Methods and Algorithms*. Wiley Interscience, Hoboken (2004)

- [11] Rastrigin, L.A., Erenstein, R.H.: *Method of Collective Recognition*. Energoizdat, Moscow (1981)
- [12] Smits, P.: Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection. *IEEE Trans. on Geoscience and Remote Sensing* 40, 717–725 (2002)
- [13] Woloszynski, T., Kurzynski, M.: On a new measure of classifier competence in the feature space. In: Kurzynski, M., Wozniak, M. (eds.) *Computer Recognition Systems 3. Advances in Intelligent and Soft Computing*, vol. 57, pp. 285–292. Springer, Heidelberg (2009)
- [14] Woloszynski, T., Kurzynski, M.: On a new measure of classifier competence applied to the design of multiclassifier systems. In: Foggia, P., Sansone, C., Vento, M. (eds.) *ICIAP 2009. LNCS*, vol. 5716, pp. 995–1004. Springer, Heidelberg (2009)
- [15] Woloszynski, M., Kurzynski, M.: A measure of competence based on randomized reference classifier for dynamic ensemble selection. In: *20th Int. Conf. on Pattern Recognition*, pp. 4194–4197. IEEE Computer Press, Istanbul (2010)
- [16] Woods, K., Kegelmeyer, W., Bowyer, W.: Combination of multiple classifiers using local accuracy estimates. *IEEE Trans. on Pattern Analysis and Machine Learning* 19, 405–410 (1997)

Privacy Preserving Models of k -NN Algorithm

Bartosz Krawczyk and Michal Wozniak

Abstract. The paper focuses on the problem of privacy preserving for classification task. This issue is quite an important subject for the machine learning approach based on distributed databases. On the basis of the study of available works devoted to privacy we propose its new definition and its taxonomy. We use this taxonomy to create several modifications of k -nearest neighbors classifier which are consistent with the proposed privacy levels. Their computational complexity are evaluated on the basis of computer experiments.

Keywords: privacy preserving data mining, distributed data mining, pattern recognition, k -NN, database security.

1 Introduction

Nowadays many enterprisers have collected huge amounts of valuable data, but their manual analysis is virtually impossible. Ganz et al. [14] report that in 2012 2.5 trillion gigabytes will be produced and storage in the computer systems. The market-leading companies realize that smart analytics tools which are capable to interpret collected data could lead to business success. Therefore they desire to exploit strength of machine learning techniques like pattern recognition methods to extract hidden, valuable knowledge from the huge, usually distributed databases. The growing amount of mentioned above distributed data collections has caused a rapid development of distributed data mining methods which are able to deal with collected data effectively.

Designing classifiers for very large databases is an attractive analytics tool [13, 17]. They are applied to the many practical areas like credit approval, prediction of customer behavior, fraud detection, designing IPS/IDS, medical diagnosis,

Bartosz Krawczyk · Michal Wozniak

Wroclaw University of Technology,

Department of Systems and Computer Networks,

Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

e-mail: bartosz.krawczyk,michal.wozniak@pwr.wroc.pl

to name only a few. Numerous approaches have been proposed to construct efficient pattern recognition methods [2]. The k -nearest neighbor rule [11], is one of the most fundamental and simple implemented nonparametric classification algorithms. The minimal distance classification is attractive from the theoretical and the practical point of view, because it is the recommended approach for discriminant analysis when the knowledge about probability densities is insufficient. However, its theoretical properties guaranteeing that for all distributions its probability of error is bounded above by twice the Bayes probability of error [9]. The naive implementation of this rule does not have any learning phase.

Let us revert to the subject of distributed data analysis. Unfortunately, we should remember that we can find legal or commercial limitations which do not allow sharing data between databases. Therefore developing privacy preserving versions of well known data analysis techniques is the focus of intense research. The aim of privacy preserving is to offer possibility of distributed data analysis with a minimum of information disclosure. The main objectives of this work are to present an original taxonomy of privacy preserving machine learning methods and to propose several modifications of k -NN classifier which allows to keep established level of privacy.

The content of the work is as follows. In the next section a pattern recognition task and k -NN algorithm are presented shortly. Then chosen topics related to privacy preserving classification are discussed and on this basis an original taxonomy of privacy preserving is introduced. Several modifications of k -NN classifier proposed on the basis of the taxonomy are described in section 4. Then the results of computer experiments which estimate computational cost of proposed algorithms are discussed. The last section concludes the paper.

2 Model of Pattern Recognition Task

The aim of the pattern recognition is to classify a given object to the one of the predefined categories, on the basis of observation of the features described it. We do not focus on a feature selection, which is very important for the quality of classifier, i.e. we assume that the set of features is given by an expert or chosen by a feature selection method [12]. All data concerning the object and its attributes are presented as a feature vector $x \in X$.

The pattern recognition algorithm Ψ maps the feature space X to the set of class labels M

$$\Psi : X \rightarrow M \quad (1)$$

The mapping (1) is established on the basis of examples included in a learning set or rules given by experts. The learning set consists of learning examples, i.e. observation of features described object and its correct classification.

2.1 k -Nearest Neighbors Classifier

Among different concepts of instance-based learning [11], the minimal distance classifier - k -nearest neighbors is simple and quite popular. It proposes to classify an

object by a majority vote of its k neighbors (k is a user-defined constant which value is fixed arbitrary using the prior knowledge about the task or its value is chosen on the basis of computer experiments). Usually Euclidean distance is used as the distance metric, but in cases, e.g. when discrete attributes are used, another metric should be chosen [8].

A weakness of the original k -nearest neighbors classifier is that the classes with the most frequent examples tend to dominate the prediction of the object, as they tend to come up in the k nearest neighbors when the neighbors are computed due to their large number. One way to overcome this problem is to weight the classification taking into account the distance from the test point to each of its k nearest neighbors [10].

The computational load of a minimal distance classification of a single object is $O(nd)$, where d is the dimension of feature vectors and n is the number of training samples. Therefore, the classification task takes a lot of time, particularly if n is large.

There are three main approaches which try to improve efficiency of k -NN rule.

1. The first category concerns methods for speeding-up nearest neighbor search [5, 10].
2. The second one proposes a training set consistent subset for the nearest neighbor rule [15], which determines a consistent subset of the original learning set. Angiulli et al. propose the distributed version of consistent subset learning algorithms for the nearest neighbor rule [3].
3. The third category concerns methods which focus on improving the classification accuracy or the response time through the use of partitions of original dataset for the nearest neighbor rule. Let us assume that an original database is divided into V partitions. We could propose two different models of the classification:
 - a. For each partition of datasets object is classified independently and the final decision is made on the basis of aforementioned V decisions using a fusion method like majority voting or more sophisticated fuser model [20].
 - b. For each partition k -nearest neighbors are found independently and finally k -nearest neighbors are selected among $k * V$ candidates coming from subsets of original dataset. Let us note that additional computations for making final decision (besides computations connected with establishing k -nearest neighbors in each partition) should be done.

In the remaining part of paper we will focus on the last category, but we have to stress flexibility of the proposition 3a. It could be used for the case that database is partitioned from privacy reasons as well because this method does not need to known feature values of the nearest neighbors.

3 Privacy Preserving

A privacy is a crucial problem in practical computer applications [6] and breaking privacy principle could lead to serious legal consequences. Westin [25] proposes the following definition of privacy: "*control over personal information*". Unfortunately

it is too general to be used in practice, therefore Moor [22] notice that the concept of privacy itself is best explained in terms of restricted access, not control. Problem of individual privacy in public spaces is widely discussed, e.g. in [23].

Privacy preserving methods are strongly related to secure databases and secure operating [16, 18]. Privacy preserving data mining methods focus on hidden information related to an individual record. One of the most popular approaches uses data perturbation techniques or secure models to protect privacy of distributed databases [21]. Some general tools supporting privacy-preserving data analysis like secure sum, comparison, division, scalar product, etc [7]. They are used in implementation of privacy-preserving in data mining methods, e.g. decision tree induction [24].

There are several works related to privacy protecting data mining for horizontal and vertical partitioned databases. For classification model under consideration it is worth noting the works which deal with privacy preserving k -NN across vertically [19] and horizontally partitioned datasets [6].

Mentioned above methods use different models, e.g. Kantarcgouglu et al. use k -anonymity privacy model [7] and Chitti et al. [6] propose to randomize data through multiple rounds.

Most of the works on privacy preserving data mining show the aspect of privacy as binary one - full privacy or lack of privacy. This observation encourages us to formulate the following definition of privacy. We assume that we would like to assure privacy in a distributed environment, e.g. where data are distributed among several nodes of computer network.

Definition 1. Each of nodes has data and view. Data can be understood as direct values, statistical information or background knowledge. Each of them can be marked as private or public, depending on a database owner. View represents the direct access to the data possessed by other node or the ability to extract or estimate those data. As preserving privacy we mean that the view of other side is restricted to the data marked as public. As the lack of privacy we mean that the view of other node allows access to private part of the database, thus creating threat to its owner.

According to formulated above definition we propose the following privacy taxonomy.

- **Level 0:** No privacy - all data are public and available.
- **Level 1:** Local privacy - individual data values are protected, but statistical data, such as histograms are public.
- **Level 2:** Statistical privacy - access to individual values is possible, but extracting statistical information is denied.
- **Level 3:** Global privacy - both direct access to individual values and statistical information are protected.
- **Level 4:** Classification step privacy - feature names used by classifier are public, but their values are private.
- **Level 5:** Classifier privacy - feature names and their values used by classifier are private. Only the result of data mining process is known to all sides.
- **Level 6:** Full privacy - all the data are restricted.

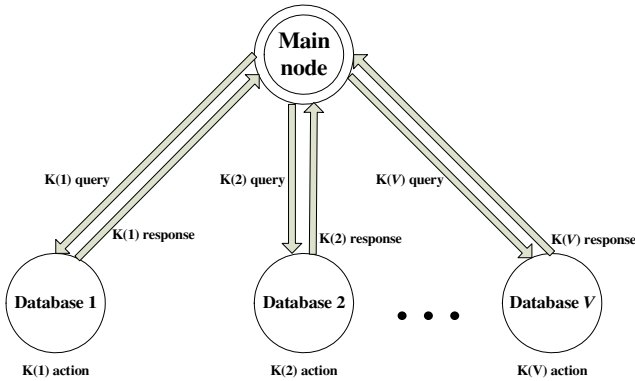


Fig. 1 Schema of described distributed data mining process with one main node and V databases.

4 Algorithmic Approach to Privacy Preserving Data Mining

For purpose of our research we focus on possibilities of protecting data by mechanism that are offered by classification algorithms themselves. Usually, privacy preserving methods are based on supplementary mechanisms like Yao’s protocol [8]. Unfortunately they usually require extra remarkable computational cost, therefore solutions that use mechanisms embedded in classification process itself are more interesting.

The k -nearest neighbour classifier offers interesting possibilities for task of privacy protecting data mining. It is worth noticing that even in its basic form this algorithm primitively conceals data itself, because the object is not used itself in the classification step, but only the value of its distance to classified pattern. Therefore we can find the hypersphere on which this object is located, but we cannot directly point at it.

We assume that in described process we use V different databases (partitions, sides, nodes) and one main node, which is responsible for conducting the algorithm. We enlist three different types of events – a query from the main node, an action taken on a database and answer which is sent from a database to the main node. Schema of this process is shown in figure 1.

On the basis of the taxonomy presented in the previous section, four different approaches for privacy preserving k -NN in distributed environment are proposed. Each of them has its own advantages and drawbacks. It is up to individual user to choose the scheme most adjusted to his needs.

4.1 Normal Querying with Full Neighbour Set

Algorithm:

V_s = set of nodes
 V = number of nodes

k = parameter of k -NN algorithm
 query all nodes from V_s for k -nearest neighbours
 main node sorts received objects
 best k objects are chosen

Benefits: Each of the nodes sends the same number of objects, therefore none of them has the knowledge which of them are used (privacy level 2 achieved).

Drawbacks: Each of the nodes sends full set of k objects. Therefore in more complex problems it reveals some direct information.

4.2 *Ranked Querying*

Algorithm:

V_s = set of nodes
 V = number of nodes
 k = parameter of k -NN algorithm
 $t = 0$
 n = specified by user, $n \ll V$
 repeat
 query all nodes from V_s for nearest neighbour
 remove n nodes with worst response from V_s set
 $t = t + 1$
 until $t == k$ or no better solution found

Benefits: Number of queries is reduced. Databases reveal significantly less information, because some of them are eliminated from querying during the procedure (privacy level 3 achieved).

Drawbacks: Database that was eliminated in beginning is informed that its object are not participating in classification. Database that was asked frequently may assume that its objects play a major role in decision process.

4.3 *Set of the Best Objects*

Algorithm:

V_s = set of nodes
 V = number of nodes
 k = parameter of k -NN algorithm
 generate set of k random objects in the main node
 denote them as the set of best objects
 repeat
 choose random node from V_s set
 send set of best objects to the chosen node
 compare objects in the node with received set


```

    replace all worse objects
    remove this node from Vs set
    return the set to the main node
    denote it as the set of best objects
    V = V - 1
until V == 0

```

Benefits: Due to random set generation in the first step none of the nodes know if they are asked as the first or the last one. Therefore no database can assume the information that the package of objects they send to the main node is the final set in the classification process. Also they are unable to identify to which databases the objects from input set belongs. Each node reveals only the objects that are sent to the main node. If database does not have better objects, none of them are revealed and it is impossible to point out which database does not return any new objects. (privacy 4 are achieved).

Drawbacks: This approach assumes that each of the nodes must compare on its side the whole received set with objects stored in database. Therefore the computational complexity is higher than in the previous methods.

4.4 Step Query with Increasing Neighbour Set

Algorithm:

```

Vs = set of nodes
V = number of nodes
k = parameter of  $k$ -NN algorithm
z = query parameter
generate set of k random objects in the main node
denote them as the set of best objects
repeat
    choose random node from Vs set
    z = k
    repeat
        send object number z from the best set to this node
        compare it with objects in the node
        if there is a better one send it to the main node
        z = z - 1
    until no better objects found or z = 0
    main node sorts the set of recieved objects
    denote them as the set of best objects
    remove this node from Vs set
    V = V - 1
until V == 0

```

Benefits: Due to random set generation in first step none of the nodes knows if it is asked as the first or the last one. Therefore no database can assume the information it sends to the main node is the final set in the classification process. Also they are unable to identify to which databases the objects from input set belong to. Each node reveals only the objects that are sent to the main node. If database does not have better objects, none of them are revealed and it is impossible to point which database does not return any new objects. Additionally number of queries is significantly reduced if tested database can not return better objects then the query ends. Each node does not receive the tested set but only one object from it at the time. If there are no better objects in database it will not receive more objects, thus concealing their values (privacy level 4 achieved). Additionally lowest number of objects is revealed compared to other method.

Drawbacks: This approach assumes that each of the nodes must compare on it's side received objects with objects stored in database. Therefore the computational complexity is the highest, compared to the previous presented methods. Additionally if a query from main ends quickly, this side may assume that its objects will not participate in the classification process.

5 Experimental Evaluation

Usually the three performance metrics for privacy preserving classification algorithm are used [6]:

1. Accuracy - which evaluates what is the accuracy lost of privacy preserving method compared with the original algorithm.
2. Efficiency - which evaluates what is the computational and memory complexity increases of privacy preserving modification compared with the original algorithm.
3. Privacy - which estimate how much information are shared between the nodes during privacy preserving modification.

Let us notice that our propositions are as accurate as the original k -NN and each of them assures the appropriate privacy level what is discussed in the previous section. Therefore the main goal of the experiments is to evaluate the computational complexity (efficiency) of proposed methods. All experiments were carried out on handwritten letter recognition dataset from UCI [4]. It consists of 20000 objects described by the sixteen features. Tests were carried out for different values of k parameter, number of nodes V , and number of features. We had assumed that 20000 is permanent number of objects and they were divided equally between the V nodes.

All experiments were carried out on Intel Core Duo T5800 2,0 GHz CPU with 3 GB RAM memory in R environment, with k -NN algorithm taken from dedicated package, thus ensuring that results achieved the best possible efficiency and that performance was not decreased by a bad implementation.

Figures from 2 to 4 show the execution time of proposed k -NN modifications.

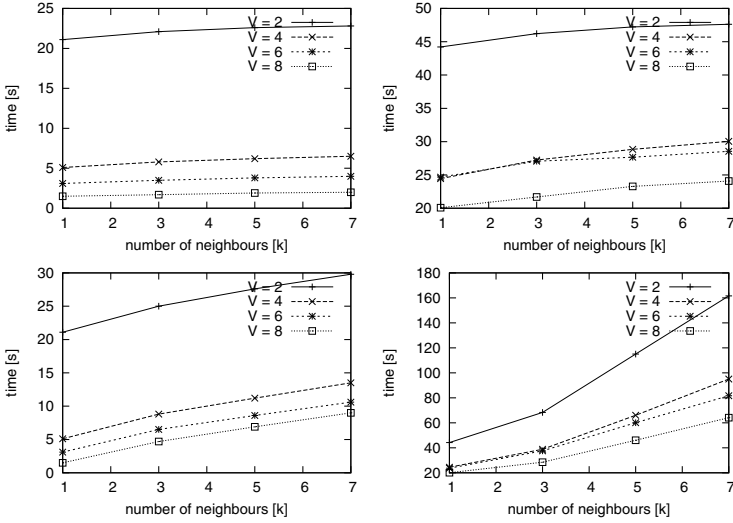


Fig. 2 Time complexity for objects with 4 features.

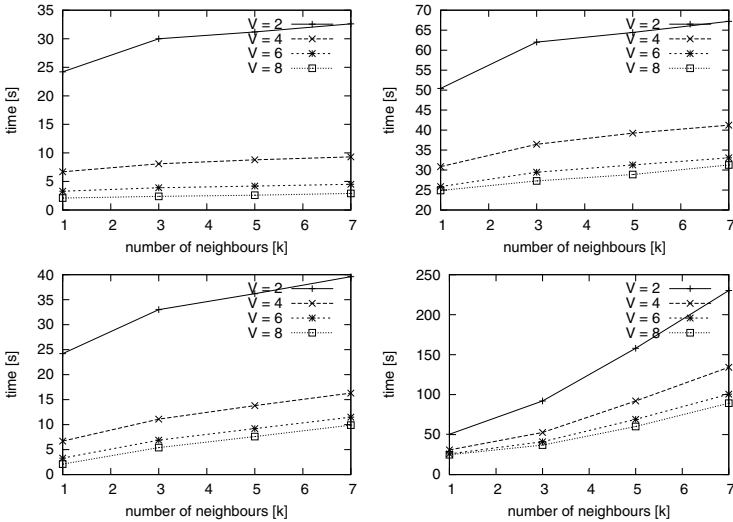


Fig. 3 Time complexity for objects with 8 features.

The results of experiments did not surprise us, because we expected that higher privacy level requires higher computational cost, but it is worth noting that the additional computational cost is lower than usually reported in literature. Of course we should remember that the decision about choosing an appropriate privacy level should be made by user.

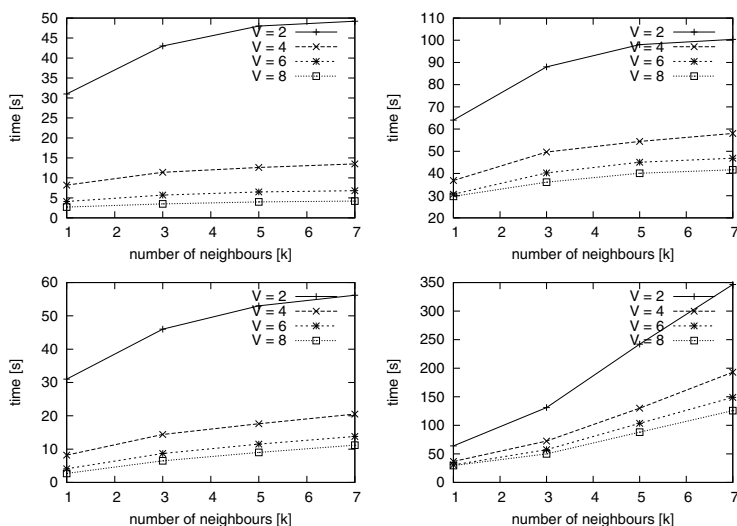


Fig. 4 Time complexity for objects with 16 features.

6 Conclusions

The paper dealt with a problem of privacy preserving for pattern recognition task. We proposed new definition of privacy and the original taxonomy of privacy preserving algorithms was formulated. Classical binary representation of privacy have strong limitations and leave a little choice for users. Therefore by splitting idea of privacy into levels we introduced a flexible framework that can be adjusted to personal needs and offers balance between safety and computational cost. We discussed four modification of the k -NN algorithm which take mentioned above privacy taxonomy into consideration. Our propositions were evaluated on the basis of computer experiments which were carried out on the basis of huge benchmark dataset. What is important, we showed that a good level of security can be achieved without using additional time consuming algorithms. We believe that proposed concept can be useful during real projects of computer recognition system which could use distributed datasets and should respect privacy aspect.

Acknowledgement. This work is supported by The Ministry of Science and Higher Education under the grant which is realizing in years 2010-2013.

References

- [1] Aha, D.W., Kibler, D., Albert, M.: k Instance-Based Learning Algorithms. *Machine Learning* 6, 37–66 (1991)
- [2] Alpaydin, E.: *Introduction to Machine Learning*, 2nd edn. The MIT Press, London (2010)

- [3] Angiulli, F., Folino, G.: Distributed Nearest Neighbor Based Condensation of Very Large Datasets. *IEEE Transactions on Knowledge and Data Engineering* 19(12), 1593–1606 (2007)
- [4] Asuncion, A., Newman, D.J.: UCI Machine Learning Repository. School of Information and Computer Science, University of California, Irvine (2007), <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [5] Beygelzimer, A., Kakade, S., Langford, J.: Cover trees for nearest neighbor. In: Proc. of the 23rd International Conference on Machine Learning, Pittsburgh, PA, pp. 97–104 (2006)
- [6] Chitti, S., Liu, L., Xiong, L.: Mining Multiple Private Databases using Privacy Preserving k NN Classifier, Technical Reports TR-2006-008, Emory University (2006)
- [7] Clifton, C., Kantarcioglu, M., Vaidya, J., Lin, X., Zhu, M.Y.: Tools for privacy preserving data mining. In: *SIGKDD Explorations*, pp. 28–34 (2002)
- [8] Cost, S., Salzberg, S.: A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features. *Machine Learning* 10(1), 57–78 (1993)
- [9] Cover, T.M., Hart, P.E.: Nearest neighbor pattern classification. *IEEE Trans. on Inform. Theory* 13(1), 21–27 (1967)
- [10] Dasarathy, B.: Nearest Neighbor (NN) Norms NN Pattern Classification Techniques. *IEEE Computer Society Press, Los Alamitos* (1991)
- [11] Devroye, L.: On the inequality of cover and hart in nearest neighbor discrimination. *IEEE Trans. on Pat. Anal. and Mach. Intel.* 3, 75–78 (1981)
- [12] Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley Interscience, New York (2001)
- [13] Freitas, A.A., Lavington, S.H.: *Mining Very Large Databases with Parallel Processing*. Kluwer Academic Publishers, Boston (1998)
- [14] Gantz, J., Reinsel, D.: *As the Economy Contracts, the Digital Universe Expands*. IDC Multimedia Whitepaper (2009)
- [15] Hart, P.E.: The condensed nearest neighbor rule. *IEEE Trans. on Inform. Th.* 14(3), 515–516 (1968)
- [16] Hacigumus, H., Iyer, B., Li, C., Mehrotra, S.: Executing sql over encrypted data in the database service provider model. In: *ACM SIGMOD Conference* (2002)
- [17] Han, J.: *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publ. Inc., San Francisco (2005)
- [18] Jajodia, S., Sandhu, R.: Toward a multilevel secure relational data model. In: *ACM SIGMOD Conference* (1991)
- [19] Kantarcioglu, M., Clifton, C.: Privacy preserving k -nn classifier. In: *IEEE International Conference on Data Engineering, ICDE* (2005)
- [20] Kuncheva, L.I.: *Combining pattern classifiers: Methods and algorithms*. Wiley-Interscience, New Jersey (2004)
- [21] Lindell, Y., Pinkas, B.: Privacy Preserving Data Mining. *Journal of Cryptology* 15(3), 177–206 (2004)
- [22] Moor, J.H.: The future of computer ethics: You ain't seen nothing yet! In: *Ethics and Information Technology*, vol. 3, pp. 89–91. Kluwer Academic Publishers, Dordrecht (2001)
- [23] Nissenbaum, H.: Can we Protect Privacy in Public? In: *Computer Ethics Philosophical Enquiry ACM/SIGCAS Conference*, Rotterdam, The Netherlands (1997)
- [24] Teng, Z., Du, W.: A hybrid multi-group privacy-preserving approach for building decision trees. In: *Proceedings of the 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining PAKDD 2006*, pp. 296–307 (2006)
- [25] Westin, A.F.: *Privacy and Freedom*. The Bodley Head Ltd (1970)

Some Properties of Binary Classifier with Fuzzy-Valued Loss Function

Robert Burduk

Abstract. In this paper we present some properties of binary classifier with fuzzy-valued loss function. The loss function in our case is dependent on the stage of the decision tree or depends on the node of the decision tree. The decision rules of a two-stage binary classifier minimize the mean risk, that is the mean value of the fuzzy loss function. In the paper the effect of a loss function on the value of the separation point of decision regions is presented. In this paper we will not go to study the impact of the choice of ranking of fuzzy numbers method on the results of the classification.

1 Introduction

In many decision problems the evaluation of utility can be quite imprecise [3]. 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. For this reason, studies thus far have previously described decision problems in which values assessed as the consequences of decision are assumed to be fuzzy [2], [6], [9]. Those papers describe only single-stage decision problems. This paper deals with a recognition problem, where - assuming a probabilistic model with full information - values of a loss function are assumed to be fuzzy numbers. We will also consider the so-called Bayesian hierarchical classifier [7]. In this recognition problem the decision as to the membership of an object in a given class is not a single activity, but is a result of a more or less complex decision making process. This model was formulated so that, on the one hand, the existence of fuzzy loss function representing the preference pattern of the decision

Robert Burduk
Department of Systems and Computer Networks,
Wroclaw University of Technology,
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
e-mail: robert.burduk@pwr.wroc.pl

maker can be established; whilst, on the other hand, a priori probabilities of classes and class-conditional probability density functions can be given.

In the paper [8] 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) of 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.

The contents of the work are as follows: Section 2 introduces the necessary background and describes the Bayes hierarchical classifier. In section 3, the recognition algorithms with fuzzy valued loss function are presented. Those values depend either on the stage of the decision tree on which an error has occurred or depend on the node of the decision tree at which misclassification has been made. In section 4, we compared those two cases of fuzzy-valued loss functions. The effect of a loss function on the value of the separation point of decision regions is presented.

2 Decision Problem Statement

Let us consider a pattern recognition problem in which the number of classes equal 4. Let us assume that the classes are organized in a two-stage binary decision tree. Let us number all the nodes of the constructed decision-tree with consecutive numbers of $0, 1, 2, \dots$, reserving 0 for the root-node and let us assign numbers of classes from the $\mathcal{M} = \{1, 2, 3, 4\}$ set to terminal nodes so that each one of them can be labelled with the number of the class connected with that node. This allows us to introduce the following notation:

- $\mathcal{M}(n)$ – the set of numbers of nodes, which distance from the root is $n, n = 0, 1, 2$.
In particular $\mathcal{M}(0) = \{0\}, \mathcal{M}(N) = \mathcal{M}$,
- $\overline{\mathcal{M}} = \bigcup_{n=0}^{N-1} \mathcal{M}(n)$ – the set of interior node numbers (non terminal),
- $\mathcal{M}_i \subseteq \mathcal{M}(N)$ – the set of class labels attainable from the i -th node ($i \in \overline{\mathcal{M}}$),
- \mathcal{M}^i – the set of numbers of immediate descendant nodes ($i \in \overline{\mathcal{M}}$),
- m_i – number of direct predecessor of the i -th node ($i \neq 0$).

We will continue to adopt the probabilistic model of the recognition problem, i.e. we will assume that the class label of the pattern being recognized $j_N \in \mathcal{M}(N)$ and its observed features x are realizations of a couple of random variables J_N and X . The complete probabilistic information denotes the knowledge of a priori probabilities of classes:

$$p(j_N) = P(J_N = j_N), \quad j_N \in \mathcal{M}(N) \quad (1)$$

and class-conditional probability density functions:

$$f_{j_N}(x) = f(x/j_N), \quad x \in X, \quad j_N \in \mathcal{M}(N). \tag{2}$$

Let

$$x_i \in X_i \subseteq R^{d_i}, \quad d_i \leq d, \quad i \in \mathcal{M} \tag{3}$$

denote vector of features used at the i -th node, which have been selected from the vector x .

Our aim is now to calculate the so-called multistage recognition strategy $\pi_N = \{\Psi_i\}_{i \in \overline{\mathcal{M}}}$, that is the set of recognition algorithms in the form:

$$\Psi_i : X_i \rightarrow \mathcal{M}^i, \quad i \in \overline{\mathcal{M}}. \tag{4}$$

Formula (4) is a decision rule (recognition algorithm) used at the i -th node which maps observation subspace to the set of immediate descendant nodes of the i -th node. Equivalently, decision rule (4) partitions observation subspace X_i into disjoint decision regions $D_{x_i}^k$, $k \in \mathcal{M}^i$, such that observation x_i is allocated to the node k if $k_i \in D_{x_i}^k$, namely:

$$D_{x_i}^k = \{x_i \in X_i : \Psi_i(x_i) = k\}, \quad k \in \mathcal{M}^i, \quad i \in \overline{\mathcal{M}}. \tag{5}$$

Let $\tilde{L}(i_N, j_N)$ denote the fuzzy loss incurred when object of the class j_N is assigned to the class i_N ($i_N, j_N \in \mathcal{M}(N)$). Our aim is to minimize the mean risk, that is the mean value of the fuzzy loss function:

$$\tilde{R}^*(\pi_N^*) = \min_{\Psi_n, \Psi_{N-1}} \tilde{R}(\pi_N) = \min_{\Psi_n, \Psi_{N-1}} \tilde{E}[L(I_N, J_N)]. \tag{6}$$

We will refer to the π_N^* strategy as the globally optimal N -stage recognition strategy.

$\tilde{R}^*(\pi_N^*)$ is a fuzzy-valued function on \mathbb{R} , taking on values of the set $\mathcal{F}_c(\mathbb{R})$ (set of normalized convex fuzzy sets on \mathbb{R} whose level sets and the closed convex hull of the support are compact).

In order to rank fuzzy mean values, we had selected the subjective method defined by Campos and González [5]. This method is based on an λ -average value of a fuzzy number, which is defined for $\tilde{A} \in \mathcal{F}_c(\mathbb{R})$ as a real number given by

$$V_S^\lambda(\tilde{A}) = \int_0^1 [\lambda a_{\alpha 2} + (1 - \lambda)a_{\alpha 1}] dS(\alpha) \tag{7}$$

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 λ parameter is a subjective degree of optimism-pessimism. In context of a loss, $\lambda = 0$ reflects the highest rate of optimism, while $\lambda = 1$ reflects the highest rate of pessimism. Then, the λ -ranking method used for comparing fuzzy numbers in $\mathcal{F}_c(\mathbb{R})$ is given by

$$\tilde{A} \succeq \tilde{B} \Leftrightarrow V_S^\lambda(\tilde{A}) \geq V_S^\lambda(\tilde{B}). \tag{8}$$

This λ -average extends some well-known ranking functions [1], [12]. One of the most relevant characteristic of the ranking method based on the function V_S^λ is its

feasibility, which is due to the following reason: when we apply V_S^λ on the fuzzy expected value of an integrably bounded fuzzy random variable the computation of this fuzzy expected value is not required. The λ -average value of a fuzzy expected value is reduced to the expected value of a classic random variable, namely, the composition of V_S^λ and the fuzzy variable.

In addition to the described method many methods of comparison and ranking of fuzzy numbers have previously been proposed in the literature. Some of them had been reviewed and compared in [4], [10], [11]. In this paper we are not going to study the impact of the choice of ranking of fuzzy numbers method on the results of the classification.

3 The Recognition Algorithm

3.1 Stage-Dependent Fuzzy Loss Function

Let $\tilde{L}(i_N, j_N)$ denote the fuzzy loss incurred when object of the class j_N is assigned to the class i_N ($i_N, j_N \in \mathcal{M}(N)$). Let us assume the following:

$$\tilde{L}(i_N, j_N) = \tilde{L}_{d(w)}^S \tag{9}$$

where w is the first common predecessor of the nodes i_N and j_N . The fuzzy loss function define as above means that the loss depends on the stage at which misclassification is made.

By putting (9) into (6), we obtain the optimal (Bayes) strategy whose decision rules at the first stage are as follows:

$$\begin{aligned} \Psi_{i_n}^*(x_{i_n}) &= i_{n+1}, \\ (\tilde{L}_0^S - \tilde{L}_1^S)p(i_{n+1})f_{i_{n+1}}(x_{i_n}) + \\ + \tilde{L}_1^S \sum_{j_N \in \mathcal{M}^{j_N-1}} [q^*(j_N/i_{n+1}, j_N)p(j_N)f_{j_N}(x_{i_n})] &= \\ = \max_{k \in \mathcal{M}^{i_n}} \left\{ (\tilde{L}_0^S - \tilde{L}_1^S)p(k)f_k(x_{i_n}) + \right. \\ \left. + \tilde{L}_1^S \sum_{j_N \in \mathcal{M}^k} [q^*(j_N/k, j_N)p(j_N)f_{j_N}(x_{i_n})] \right\} \end{aligned} \tag{10}$$

where $q^*(j_N/i_{n+1}, j_N)$ denotes the probability of accurate classification of the object of the class j_N at the second stage using π_N^* strategy rules, on condition that at the first stage the i_{n+1} decision has been made.

Decision rules at the second stage of classification are like for single-stage classifier with zero-one loss function.

3.2 Node-Dependent Fuzzy Loss Function

Let us assume now:

$$\tilde{L}(i_N, j_N) = \tilde{L}_w \tag{11}$$

where w is the first common predecessor of the nodes i_N and j_N . The fuzzy loss function defined as above means that the loss depends on the node of the decision tree at which misclassification has been made. The difference between stage-dependent fuzzy loss function and node-dependent fuzzy loss function for the two-stage binary classifier is presented in Fig. 1

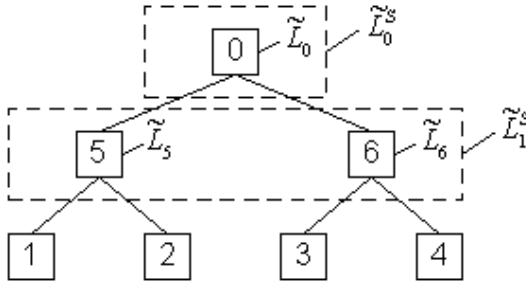


Fig. 1 Interpretation of stage-dependent and node-dependent fuzzy loss function

By putting (11) into (6), we obtain the optimal (Bayes) strategy whose decision rules at the first stage are as follows:

$$\begin{aligned}
 & \Psi_{i_n}^*(x_{i_n}) = i_{n+1}, \\
 & (\tilde{L}_0 - \tilde{L}_{i_{n+1}})p(i_{n+1})f_{i_{n+1}}(x_{i_n}) + \\
 & + \tilde{L}_{i_{n+1}} \sum_{j_N \in \mathcal{M}^{j_N-1}} [q^*(j_N/i_{n+1}, j_N)p(j_N)f_{j_N}(x_{i_n})] = \\
 & = \max_{k \in \mathcal{M}^{i_n}} \left\{ (\tilde{L}_0^S - \tilde{L}_k) p(k) f_k(x_{i_n}) + \right. \\
 & \left. + \tilde{L}_k \sum_{j_N \in \mathcal{M}^k} [q^*(j_N/k, j_N)p(j_N)f_{j_N}(x_{i_n})] \right\}
 \end{aligned} \tag{12}$$

where $q^*(j_N/i_{n+1}, j_N)$ denotes the probability of accurate classification of the object of the class j_N at the second stage using π_N^* strategy rules, on condition that at the first stage the i_{n+1} decision has been made.

Decision rules at the second stage of classification are like for single-stage classifier with zero-one loss function.

4 Some Properties of the Binary Classifier

Let us consider the two-stage binary classifier like the one presented in Fig. 1. We assume now that we have conditional independent features: $f_5(x_0) = f_1(x_0) = f_2(x_0)$ and $f_6(x_0) = f_3(x_0) = f_4(x_0)$. Additionally we assume that we have equal a priori probabilities of classes: $p(1) = p(2) = p(3) = p(4)$. For those assumptions the recognition algorithm at the first stage of classification for the stage-dependent fuzzy loss function (9) is following:

$$\Psi_0^*(x_0) = i_{n+1}, \text{ when}$$

$$i_{n+1} = \max_{k \in \mathcal{M}^0} \left\{ (\tilde{L}_0^S - \tilde{L}_1^S) p(k) f_k(x_0) + \right. \\ \left. + \tilde{L}_1^S p(k) f_k(x_0) \sum_{j_N \in \mathcal{M}^k} q^*(j_N/k, j_N) \right\}. \quad (13)$$

Knowing that $p(5) = p(1) + p(2)$, $p(6) = p(3) + p(4)$ and $k \in 5, 6$ the point x_0^* (the separation point of decision regions at the first stage) is calculated from the equation:

$$f_5(x_0)(\tilde{L}_0^S - \tilde{L}_1^S + \tilde{L}_1^S(q^*(1/5, 1) + q^*(1/5, 1))) = \\ f_6(x_0)(\tilde{L}_0^S - \tilde{L}_1^S + \tilde{L}_1^S(q^*(3/6, 3) + q^*(4/6, 4))).$$

Comparing the obtained results with the algorithm for the zero-one loss function and locally optimal strategy it can be easily observed that the algorithm prefers the node k of tree for which the sum of probabilities of correct classification $q^*(j_N/k, j_N)$ is greater. Preference is defined as an increase in decision regions, or equivalently, the shift of the separation point of decision regions x_0^* . Values of loss functions effect the value of a point, but preferred is always going to be the node for which greater is the value $\sum_{j_N \in \mathcal{M}^k} q^*(j_N/k, j_N)$.

It is different in the case of the node-dependent fuzzy loss function. For the same assumptions, the point x_0^* is calculated from the equation:

$$f_5(x_0)(\tilde{L}_0 - \tilde{L}_5 + \tilde{L}_5(q^*(1/5, 1) + q^*(1/5, 1))) = \\ f_6(x_0)(\tilde{L}_0 - \tilde{L}_6 + \tilde{L}_6(q^*(3/6, 3) + q^*(4/6, 4))).$$

Which node of the decision tree is preferred depends not only on the sum of probabilities of correct classification but also on the values of the loss function. Preferred is the decision tree node for which the assigned loss function is to smaller. Thus, in this case it may be preferable, each of the nodes, not just one. Which one depends on the values of loss functions, and probabilities of correct classification. For two of those loss functions we will receive the same point x_0^* value, if the assumed the node-dependent fuzzy loss functions satisfy the condition $\tilde{L}_6 = \frac{\tilde{L}_5(q^*(1/5, 1) + q^*(1/5, 1) - 1)}{q^*(3/6, 3) + q^*(4/6, 4) - 1}$. The separation point of decision regions x_0^* also depends on the method of comparing fuzzy numbers or value of parameter λ of the method described in this paper.

5 Conclusions

In the paper we have presented the multistage recognition problem in which consequences of a wrong decision are fuzzy-valued. We considered two cases of loss function. In the first one, the loss function is dependent on the stage of the decision tree, in the second it depends on the node of the decision tree. For the two-stage Bayes classifier with full probabilistic information we compared both the cases of loss function. In particular, the effect of a loss function on the value of the

separation point of decision regions was presented. In a future work the impact of the choice of ranking of fuzzy numbers method on the value of the separation point can be considered.

Acknowledgements. This work is supported by The Polish Ministry of Science and Higher Education under the grant which is being realized in years 2011-2013.

References

- [1] Adamo, J.M.: Fuzzy Decision Trees. *Fuzzy Sets and Sys.* 4, 207–219 (1980)
- [2] Baas, S., Kwakernaak, H.: Rating and Ranking of Multi-Aspect Alternatives Using Fuzzy Sets. *Automatica* 13, 47–58 (1997)
- [3] Berger, J.: *Statistical Decision Theory and Bayesian Analysis*. Springer, New York (1993)
- [4] Bortolan, G., Degani, R.: A review of some methods for ranking fuzzy subsets. *Fuzzy Sets and Systems* 80, 167–176 (1985)
- [5] Campos, L., González, A.: A Subjective Approach for Ranking Fuzzy Numbers. *Fuzzy Sets and Sys.* 29, 145–153 (1989)
- [6] Jain, R.: Decision-Making in the Presence of Fuzzy Variables. *IEEE Trans. Sys. Man and Cyber.* 6, 698–703 (1976)
- [7] Kurzyński, M.: Decision Rules for a Hierarchical Classifier. *Pat. Rec. Let.* 1, 305–310 (1983)
- [8] Kurzyński, M.: On the Multistage Bayes Classifier. *Pat. Rec.* 21, 355–365 (1988)
- [9] Viertl, R.: *Statistical Methods for Non-Precise Data*. CRC Press, Boca Raton (1996)
- [10] Wang, X., Kerre, E.E.: Reasonable properties for the ordering of fuzzy quantities (I). *Fuzzy Sets and Systems* 118, 375–385 (2001)
- [11] Wang, X., Kerre, E.E.: Reasonable properties for the ordering of fuzzy quantities (II). *Fuzzy Sets and Systems* 118, 387–405 (2001)
- [12] Yager, R.: A Procedure for Ordering Fuzzy Subsets of the Unit Interval. *Inf. Scien.* 22, 143–160 (1981)

Part III
IMAGE PROCESSING AND
COMPUTER VISION

Interactive Exploration of Image Collections

Gerald Schaefer

Abstract. With image collections, both private and commercial, ever growing, efficient and effective tools for managing these repositories are becoming increasingly important. Content-based approaches, which are based on the principle of image feature extraction and similarity calculation based on these features, seem necessary as most images are unannotated. However, typical content-based retrieval systems have shown only limited usefulness. In this paper, we present interactive image database browsing systems as an alternative to retrieval approaches. Exploiting content-based concepts, image collections can be visualised so that visually similar images are located close in the visualisation space. Once an image collection has been displayed, the user is given the possibility of interactively exploring it further through various browsing operations. After introducing basic approaches to visualising and browsing image collections, we then focus on some of the systems that we have developed in our lab for this purpose. In particular, we look in detail at the Hue Sphere Image Browser and Honeycomb Image Browser systems, both of which provide hierarchical browsing approaches that give access to large image collections in an intuitive yet efficient manner.

1 Introduction

With image collections, both private and commercial, ever growing (while personal image collections may contain hundreds or thousands of pictures, commercial image repositories can comprise millions of images [13]), efficient and effective tools for managing these repositories are becoming increasingly important. Content-based approaches [28], which are based on the principle of image feature extraction and similarity calculation based on these features, seem necessary as most images are

Gerald Schaefer
Department of Computer Science
Loughborough University
Loughborough, U.K.
e-mail: gerald.schaefer@ieee.org

unannotated. However, typical content-based retrieval systems have shown only limited usefulness and cannot be employed for generic image search, but rather only for specific specialised domains. This is partly due to the semantic gap between low-level image features and our high-level understanding of images [28], but also to other reasons such as the fact that typical retrieval systems are often based on inadequate search paradigms such as query-by-example [29]. Image retrieval through query-by-example returns, based on a query image provided by the user, those images from the database which are deemed most similar to the given query, i.e. those images whose image features best match the query. Since only a relatively small set of images (typically a few tens) is returned, most of the image database remains unknown to the user.

In this paper, we present techniques that allow full exploration of an image database. Rather than following the paradigm of direct image retrieval, image browsing systems provide an overview of a complete image repository coupled with methods to interactively navigate further through the image collection [18]. First, in Section 2, we summarise various approaches for visualising and navigating image collections [16]. In the following two sections, we then describe two of the systems we developed in our lab for interactive navigation of large image collections, namely the Hue Sphere Image Browser [23] in Section 3, and the Honeycomb Image Browser [17] in Section 4. We conclude the paper and highlight the issue of benchmarking [14] in Section 5.

2 Image Database Navigation Systems

Common tools for managing image collections typically display images in a one-dimensional linear format (usually on a grid) where only a limited number of thumbnail images are visible on screen at any one time, thus requiring the user to search back and forth through thumbnail pages to view all images. Obviously, this constitutes a time consuming, impractical and exhaustive way of searching images, especially in larger collections. Furthermore, the order in which the pictures are displayed is based on attributes like file names that often do not reflect the actual image contents and hence cannot be used to speed up the search.

A much better approach is to employ content-based image retrieval concepts [28] in order to arrange images in such a way that visually similar images are located close to each other in the display, which has been shown to decrease the time it takes to localise certain images [20]. Various approaches following this principle have been introduced in recent years [18]; in general we can divide them into mapping-based, clustering-based and graph-based image database visualisation methods [16]. Once an image collection has been visualised, interactive browsing allows full exploration of the complete image repository [15].

Mapping-based approaches are based on the idea of reducing the high-dimensional feature space into a low-dimensional visualisation space through application of a dimensionality reduction technique. The simplest of these is principal

component analysis (PCA) which uses the eigenvectors derived from the covariance matrix to plot the original data where image thumbnails are displayed at the co-ordinates obtained through projection of the original feature data in the low-dimensional space. The Personal Digital Historian project [10] uses such PCA splats in order to visualise images, while Keller *et al.* also employ a PCA visualisation to present images in a 3D interface based on texture features [8]. In contrast to PCA, multi-dimensional scaling (MDS) attempts to preserve the original relationships (i.e., distances) in the high-dimensional feature space as best possible in the low-dimensional projection and therefore works also with underlying distance measures that are different from Euclidean distances. MDS was employed by Rubner *et al.* [22] who, based on colour signatures of images, placed image thumbnails at the co-ordinates derived by the algorithm. Nguyen and Worring [12] investigated other dimensionality reduction algorithms including ISOMAP (isometric mapping), SNE (stochastic neighbour embedding) and LLE (local linear embedding) as well as some hybrid techniques for visualising image collections. A common drawback of all dimensionality reduction techniques is that they require extensive computations, in particular for large image databases. Furthermore, since images are typically placed on calculated co-ordinates, many images are occluded while others overlap each other, leading to a less intuitive browsing experience [20].

Clustering-based visualisation methods group similar images together and summarise groups through representative images. This way an initial overview even of large image sets can be gained. Krischnamachari and Abdel-Mottaleb [9] were among the first to propose clustering images by image content (based on local colour histograms). A hierarchical tree structure for browsing image databases was suggested in [3], where images are grouped in a quadtree similarity pyramid through a clustering process. A similar hierarchical browsing strategy was also pursued in [1], however in this approach the underlying tree structure can also be interactively modified by the user, while other recent systems that cluster images in a hierarchical manner for image database browsing include [4, 6, 19]. Similar to mapping-based approaches, the computational complexity involved in clustering-based approaches is usually high since clustering itself is a costly operation.

Graph-based visualisations use, as the name suggests, a graph as a visualisation (and browsing) structure, where the nodes of the graph represent images while the edges link related images (e.g. visually similar images). Dontcheva *et al.* [5] and Worring *et al.* [30] use a mass spring model to generate a visualisation based on associated keywords between images. In order to visualise this high-dimensional data in 2D, connected images are placed closer together, while unconnected images are moved further apart. Image browsing based on Pathfinder networks was introduced in [2], while Heesch and Rueger [7] proposed NN^k networks (where NN stands for nearest neighbour and k describes a set of different features) to browse through an image database.

3 Hue Sphere Image Browser

The Hue Sphere Image Browser developed in [23, 24, 25, 26] aims to provide an intuitive, effective and efficient interface for image database visualisation and navigation. In particular, we want to address the high computational demands that characterise other approaches, while integrating the advantages of both mapping- and clustering-based systems. For this, we employ simple colour-based features, namely the median colour expressed in HSV colour space of which we extract the hue and value components. While this might at first seem a bit simplistic, experiments have shown that such simple features work as well as high-dimensional descriptors describing image colour distributions [21].

Since hue is a circular quality, and the extremes of value converge to singularities (pure white respectively pure black), the visualisation space we naturally end up with is that of a spherical globe, with black and white forming the poles and hue changing in circles parallel to the equator. Since most users will be familiar with the concept of the earth globe, this provides an immediately intuitive visualisation and browsing interface as users are already aware of how to locate and find something on its surface.

The Hue Sphere Image Browser approach to visualising image databases is simple and straightforward while providing a very fast (real-time) method for accessing the image collection. Images are transformed to an HSV representation, and the median hue and value attributes calculated (we use of the median rather than the mean in order to provide some robustness with regards to the image background). From these, the co-ordinates on the globe are then determined, where hue translates to longitude and value is rescaled to match the latitude range ($V = 0.5$ corresponds to 0° latitude, i.e. a position on the equator). A thumbnail of the image is then projected onto the surface of the sphere at the calculated co-ordinates.

As mentioned above, just plotting images at calculated co-ordinates will lead to images overlapping and occluding each other. While approaches to minimise these effects have been proposed, e.g. in [10] where images are slightly moved as a result of a local optimisation problem, these constitute only a partial solution to the problem, whereas it has been shown that organisation of images on a regular grid without overlapping has a positive effect on the browsing experience [20]. In our approach, we therefore adopt a lattice structure onto which image thumbnails are placed so that they do not overlap each other. That is, we carry out the initial hue/brightness calculation and then map each image to its closest cell in the lattice.

Clearly, and in particular for larger image sets, this will mean that more than one image can be mapped to a particular cell. We consequently employ a hierarchical tree structure to address the navigation through large image collections, but also to eliminate the need for further computations. Hierarchical browsing environments such as the one described in [3] have been shown to provide an effective and efficient way of moving through large image datasets. In our approach, we make direct use of the grid mapping introduced above to build a hierarchical tree based on clustering images. However, in contrast to methods such as [3] we do not need to employ computationally expensive clustering algorithms. Rather, the resolution of the

grid layout represents the maximal number of clusters present at a given level while the grid cells directly determine which images fall into which clusters. Assigning images to clusters/grid cells can hence be performed with simple comparison operations only. This procedure is adopted at the each level of the tree hierarchy, i.e. first at the root node (the initial global display) and then for each non-empty cell again in a recursive manner, where the images of each child node are again mapped to a grid structure, until the complete tree is derived. For grid cells which contain more than one image, a representative image, selected as the image whose colour is closest to that of the centre of the cell, is used in the visualisation.

The resulting interface provides an intuitive way of browsing to the user, who can, starting from an initial display, successively select a representative image to refine the search. That image cluster is then expanded in the next level of the display where the user can again select an image group to navigate further into the image collection.

In order to reduce the number of empty grid cells, we also apply two strategies designed to fill more cells with images. The first identifies empty cells and subsequently inspects the neighbouring cells. If 3 or 4 of the 4-neighbourhood cells are occupied, a relative percentage of those images (the images closest to the borders) is moved across to fill the previously empty cell. The second filling strategy is applied to tree node cells which contain only a few images, most of which will be visually fairly similar. To avoid them from being mapped to the same cell and hence to trigger another tree level, a spreading algorithm is applied which moves images to neighbouring cells in a similar fashion as the method presented in [20].

In Figures 1 to 3 we show some screenshots of the application loaded with the MPEG-7 common colour dataset [11], an image collection of about 4500 images. A global view of the image database is given in Figure 1. As can be seen, the globe as a visualisation body provides an intuitive interface for navigation where it is clear to the user in which part to look for certain image classes.



Fig. 1 Global Hue Sphere view of the MPEG-7 dataset.



Fig. 2 Hue Sphere after after rotation and tilt operations.



Fig. 3 Hue Sphere after zoom operation.

In Figures 2 and 3 we show the results of some user interaction where the user first rotated the sphere to focus on images with a different hue followed by a tilt operation to bring up darker images resulting in the view given in Figure 2. Figure 3 then shows the result of a zoom operation where the user chose one of the images to bring up those pictures that are contained in that selected part of the tree. To aid navigation the previous hierarchy level is also displayed and the current position within those grid marked by the red dot.

Apart from the stand-alone application developed in [23], an on-line browsing tool based on the same ideas was also implemented [26] which, in addition, also offers a browsing method where the user is immersed in the sphere rather than observing it from the outside as is shown in Figure 4.

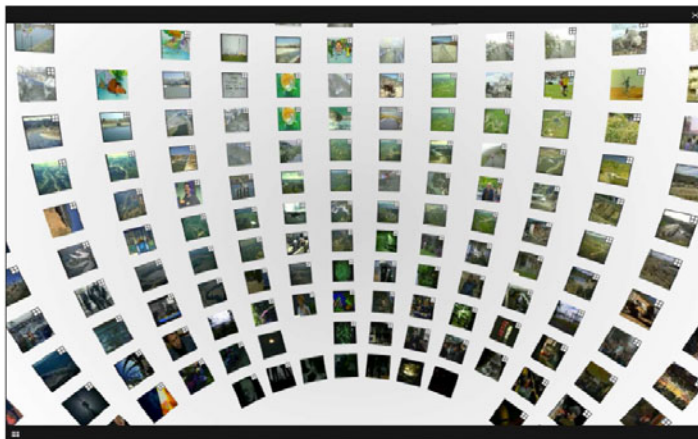


Fig. 4 Visualisation of MPEG-7 database in the online-browser [26] with the user immersed at the centre of the sphere.

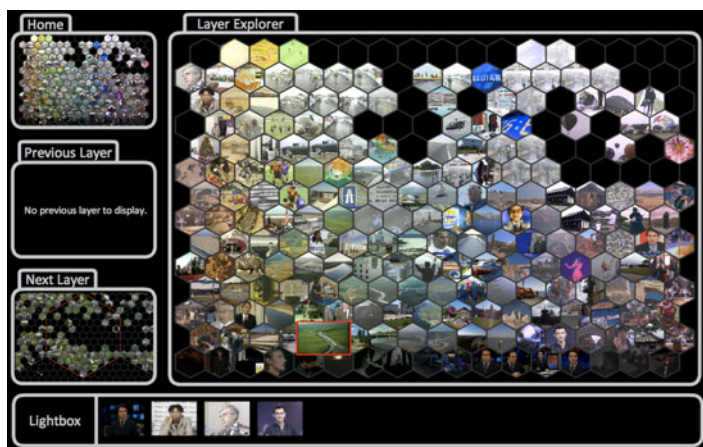


Fig. 5 MPEG-7 databased visualised in the Honeycomb Image Browser.

4 Honeycomb Image Browser

In our Honeycomb Image Browser system [17] we use many of the ideas developed for the Hue Sphere browser, namely the use of simple colour features and a regular lattice rather than utilising expensive dimensionality reduction and/or clustering techniques. However, instead of employing a spherical visualisation space, we use (as the name suggests) a hexagonal lattice. This has the advantage that, when images are organised in a space-filling arrangement, each row and column of images is visually displaced from its neighbouring rows/columns. This would not be possible

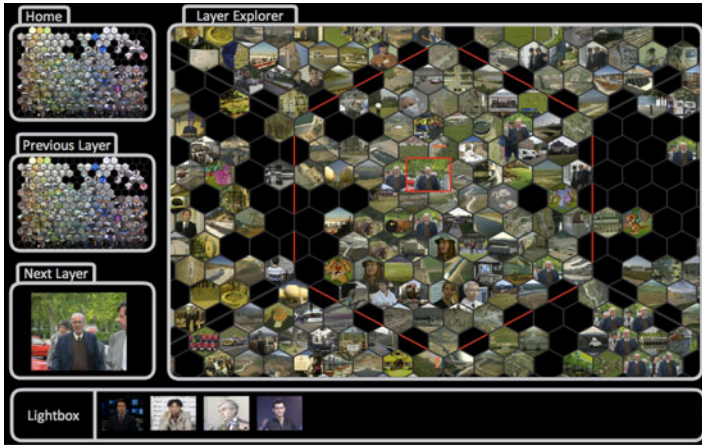


Fig. 6 Honeycomb Image Browser after navigating into the next layer of the visualisation.

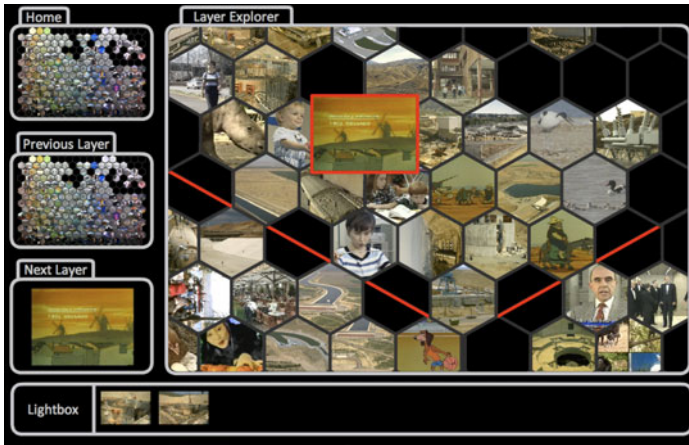


Fig. 7 Zoomed-in view of the Honeycomb Image Browser.

using a regular square grid structure as larger visual gaps are needed to delineate images clearly. The space saved as a result of using a hexagonal lattice enables larger or more images to be displayed within the visualisation. Also, on a hexagonal lattice, the six neighbours of a hexagon are equidistant from the middle cell while on a square architecture, the neighbours at the diagonal are further away than the horizontal and vertical neighbouring cells [27].

In addition, the Honeycomb Image Browser offers various other enhancements, namely the inclusion of small root and parent node views, a preview which is activated by moving the cursor over any cell, a clipboard to which the user can move images of interest, as well as continuous panning on all layers of the browsing structure. Example screenshots of the Honeycomb browser are shown in Figures 5 to 7.

5 Conclusions

Efficient and effective methods for accessing and managing large image collections are highly sought after. In this paper, we have presented image database navigation systems as an alternative to classical retrieval approaches. We have shown that intuitive and fast browsing even through large image repositories is possible through a combination of simple (colour) features and a hierarchical approach of navigation coupled with arrangement of images on a regular lattice, either on a regular grid or on a honeycomb structure.

While the usefulness of these approaches has been demonstrated, an appropriate evaluation of image browsing systems is difficult and hence often (including, admittedly, in our own work) missing as a whole or at least in part. To address this, we have set out some guidelines for benchmarking image database navigation systems [14], which include a series of tasks such as target and category searches as well as an annotation task which can be measured objectively (in terms of timing and correctness) as well as some (subjective) user feedback.

Acknowledgements. The author would like to acknowledge Simon Ruszala and William Plant who were significantly involved in the work presented in this paper.

References

- [1] Bartolini, I., Ciaccia, P., Patella, M.: Adaptively browsing image databases with PIBE. *Multimedia Tools and Applications* 31(3), 269–286 (2006)
- [2] Chen, C., Gagaudakis, G., Rosin, P.: Similarity-Based Image Browsing. In: *Int. Conference on Intelligent Information Processing*, pp. 206–213 (2000)
- [3] Chen, J.Y., Bouman, C.A., Dalton, J.C.: Hierarchical Browsing and Search of large Image Databases. *IEEE Trans. Image Processing* 9(3), 442–455 (2000)
- [4] Chen, Y., Butz, A.: PhotoSim: Tightly integrating image analysis into a photo browsing UI. In: *Int. Symposium on Smart Graphics*, pp. 224–231 (2008)
- [5] Dontcheva, M., Agrawala, M., Cohen, M.: Metadata visualization for image browsing. In: *18th Annual ACM Symposium on User Interface Software and Technology* (2005)
- [6] Gomi, A., Miyazaki, R., Itoh, T., Li, J.: CAT: A hierarchical image browser using a rectangle packing technique. In: *12th Int. Conference on Information Visualization*, pp. 82–87 (2008)
- [7] Heesch, D., Rüger, S.: NNk networks for content-based image retrieval. In: McDonald, S., Tait, J.I. (eds.) *ECIR 2004. LNCS*, vol. 2997, pp. 253–266. Springer, Heidelberg (2004)
- [8] Keller, I., Meiers, T., Ellerbrock, T., Sikora, T.: Image browsing with PCA-assisted user-interaction. In: *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pp. 102–108 (2001)
- [9] Krischnamachari, S., Abdel-Mottaleb, M.: Image browsing using hierarchical clustering. In: *IEEE Symposium on Computers and Communications*, pp. 301–307 (1999)
- [10] Moghaddam, B., Tian, Q., Lesh, N., Shen, C., Huang, T.S.: Visualization and user-modeling for browsing personal photo libraries. *Int. Journal of Computer Vision* 56(1–2), 109–130 (2004)
- [11] Moving Picture Experts Group. Description of core experiments for MPEG-7 color/texture descriptors. Technical Report ISO/IEC JTC1/SC29/WG11/ N2929 (1999)

- [12] Nguyen, G.P., Worring, M.: Interactive access to large image collections using similarity-based visualization. *Journal of Visual Languages and Computing* 19(2), 203–224 (2008)
- [13] Osman, T., Thakker, D., Schaefer, G., Lakin, P.: An integrative semantic framework for image annotation and retrieval. In: *IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 366–373 (2007)
- [14] Plant, W., Schaefer, G.: Evaluation and benchmarking of image database navigation tools. In: *Int. Conference on Image Processing, Computer Vision and Pattern Recognition*, vol. 1, pp. 248–254 (2009)
- [15] Plant, W., Schaefer, G.: Navigation and browsing of image databases. In: *Int. Conference on Soft Computing and Pattern Recognition*, pp. 750–755 (2009)
- [16] Plant, W., Schaefer, G.: Visualising image databases. In: *IEEE Int. Workshop on Multimedia Signal Processing*, pp. 1–6 (2009)
- [17] Plant, W., Schaefer, G.: Image retrieval on the honeycomb image browser. In: *17th IEEE Int. Conference on Image Processing*, pp. 3161–3164 (2010)
- [18] Plant, W., Schaefer, G.: Visualisation and browsing of image databases. In: *Multimedia Analysis, Processing and Communications. SCI*, vol. 346, Springer, Heidelberg (2010)
- [19] Platt, J., Czerwinski, M., Field, B.: PhotoTOC: automatic clustering for browsing personal photographs. Technical report, Microsoft Research (2002)
- [20] Rodden, K., Basalaj, W., Sinclair, D., Wood, K.: Evaluating a visualisation of image similarity as a tool for image browsing. In: *IEEE Symposium on Information Visualization*, pp. 36–43 (1999)
- [21] Rodden, K., Basalaj, W., Sinclair, D., Wood, K.: A comparison of measures for visualising image similarity. In: *The Challenge of Image Retrieval* (2000)
- [22] Rubner, Y., Guibas, L., Tomasi, C.: The earth mover’s distance, multi-dimensional scaling, and color-based image retrieval. In: *Image Understanding Workshop*, pp. 661–668 (1997)
- [23] Schaefer, G.: A next generation browsing environment for large image repositories. *Multimedia Tools and Applications* 47(1), 105–120 (2010)
- [24] Schaefer, G., Ruzsala, S.: Image database navigation: A globe-al approach. In: *Bebis, G., Boyle, R., Koracin, D., Parvin, B. (eds.) ISVC 2005. LNCS*, vol. 3804, pp. 279–286. Springer, Heidelberg (2005)
- [25] Schaefer, G., Ruzsala, S.: Hierarchical image database navigation on a hue sphere. In: *Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Nefian, A., Meenakshisundaram, G., Pascucci, V., Zara, J., Molineros, J., Theisel, H., Malzbender, T. (eds.) ISVC 2006. LNCS*, vol. 4292, pp. 814–823. Springer, Heidelberg (2006)
- [26] Schaefer, G., Stuttard, M.: An on-line tool for browsing large image repositories. In: *Int. Conference on Information Retrieval and Knowledge Management*, pp. 102–106 (2010)
- [27] Sheridan, P., Hintz, T., Alexander, D.: Pseudo-invariant image transformations on a hexagonal lattice. *Image and Vision Computing* 18, 907–917 (2000)
- [28] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(12), 1249–1380 (2000)
- [29] Swain, M.J., Ballard, D.H.: Color indexing. *Int. Journal of Computer Vision* 7(11), 11–32 (1991)
- [30] Worring, M., de Rooij, O., van Rijn, T.: Browsing visual collections using graphs. In: *Int. Workshop on Workshop on Multimedia Information Retrieval*, pp. 307–312 (2007)

Variable Complexity Reduction in Image Grade Decomposition

Maria Grzegorek

Abstract. The paper presents improved method of applying Grade Correspondence Cluster Analysis to image decomposition. Image pixels are set of object and can be described by a collection of values of variables. The set of pixels forms a data table which is processed by application implementing GCCA. Variables with minor complexity reveals hidden structures in more distinct manner.

1 Introduction

One of the tasks of image processing is to reveal structures in images. Some structures may be hidden as in medical imaging and only very skilled person could point these features out. Other are obvious at a glance as at photograph, however, involving very demanding task engaging context data in various applications.

In this article a next step towards application of the developed image grade decomposition is described. The used tool is statistical application GradeStat [6], which has been developed in Institute of Computer Science PAS. GradeStat makes possible analyzing data of different domains. Beyond some typical tasks it offers unique way of performing Grade Correspondence Cluster Analysis supported by extensive visualization which allows to get "deeper insight" into multivariate data sets. Wide fundamental description of grade methods is gathered in [11]. Short explanation along with economic data discussion namely climate in Polish sevicng sector, is printed in [5]. Another very short introduction to GradeStat with a theoretical example can be found in [16]. Wide and useful description how to use GradeStat in order to analyze medical and demographical data is put in [13]. Image as a source of data table analyzed with grade procedures is presented in [7] as well as the use of a grade differentiation measure between two data sets to compare variables drawn from the image [10]. Review of numerous applications is in preparation.

Maria Grzegorek
Institute of Computer Science PAS,
Ordonia 21, 01-237 Warsaw, Poland
e-mail: mary@pipan.waw.pl

There are numerous algorithms performing similar task of image segmentation, some popular are mentioned in Szeliski [15] Chapter 5. Algorithm developed by Estrada and Jepson [11] belongs to graph-based min-cut type. Pixels are nodes of graph, weights of edges are proportional to the similarity between image elements corresponding to them and build affinity matrix, entry to random walk, which results in blur kernels with a properties which allow reveal candidates to minimum cut through the graph. Shi and Malik in [14] apply disassociation measure between two groups of nodes which is normalized, e.g. sum of node connections potentially removed divided by a sum of connection of first group with all nodes plus sum of node connections potentially removed divided by a sum of connection of second group with all nodes. Second smallest of the generalized eigevalue system is a normalized cut task. Felzenszwalb and Huttenlocher [4] use measure of internal variation in region as maximum of all edge weights in minimum spanning tree and external variation measure between regions as minimum weight of edges linking these regions. Regions are merged if external measure is small enough in respect to both internal values. Comaniciu and Meer [3] apply mapping of point of data into a point in a multidimensional space of parameters. Point of data consisting for example of space coordinates, gray or colour, texture parameters and so on. In parameter space density estimate is defined, weighted mean within local neighborhood is computed and a center of local neighborhood is shifted to this mean. Point of convergence with corresponding start points determine separate region.

Grade transform groups and clusters similar pixels. Global information is preserved, local is used throughout features of neighboring pixels. So obtained regions are homogenous and present considerably basis to evolution significant segmentation.

In next Section simplified sets of variables are presented. Sets of variables are tested and illustrated with synthetic image. Section 3 describes effect of grade image decomposition and Section 4 summarizes obtained results and outlines future work.

2 Image Decomposition by GCCA

Fundamentally application GradeStat is dedicated to support data exploration. GradeStat helps to detect regularity in data table after its reordering, to find outliers and irregular objects, to detect more regular subpopulations, to divide data table into clusters.

Grade data analysis is convenient for any measurement scale because such analysis bases on dissimilarity. Therefore grade analysis performed with GradeStat can be applied to any kind of refined variables which can be defined for pixels. Two-way objects/variables table is subject to grade transformation. Proper coding conditions are easy fulfilled when pixel variable definitions are thought out.

Grade analysis involve Grade Correspondence Analysis (GCA) using grade transformation which reorders objects (pixels) and variables of the data table in such way that the table becomes more regular in regard to monotone dependence measures [11]. It results in similarity of objects (pixels) located near each other

in the reordered table and dissimilarity distant ones. Next step is objects clustering (hence Grade Correspondence Cluster Analysis – GCCA) on the aggregation rule into pixels subsets. There is also possibility to point out main trend in data to separate outliers which can be objects highly departing from this trend, for example noisy pixels. Besides numerous graphs GradeStat has an important supporting tool – overrepresentation map. This map is a graphic presentation of the probability density of grade distribution, intrinsic representation of data table in GradeStat (meaning and short interpretation can be seen in [16]). The overrepresentation map can assist in establishing the number of pixel clusters.

Pixel " p, q " belongs to row p and column q in the image. For shortening of description these indices will be omitted in following text and the term "pixel" will mean just pixel " p, q ". Pixels in the image form set of objects. Each pixel is attached by a set of variable values. Set of pixels is represented now by two dimensional data table and can be processed by grade procedures. Main task is to construct variables in such a way that cause GradeStat to cluster similar pixels, e.g. pixels with similar characteristics in the grade transformed table. Recovering spatial location of pixels in each cluster often reveals hidden structures.

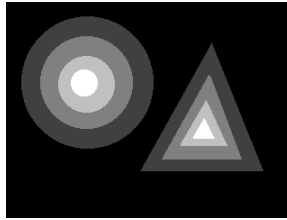


Fig. 1 Synthetic image

First obvious variable value in gray scale images is value of gray intensity g . Value of second variable may be established as a kind of gradient value, for instance Sobel operator provides gradient magnitude value m . Next values of l variables $n_i, i = 1, \dots, l$ are constructed in different way. Pixel neighborhood 3×3 is taken into consideration. In [8] variables are established on a basis of pixels magnitude values in such 8-neighborhood. Let $m_j, j = 1, \dots, k; k = 8$ are values of gradient magnitudes in 8 neighboring pixels. $t_i, i = 1, \dots, l$ is a family of thresholds values usually established as a fraction $f, f \in \langle 0, 1 \rangle$ of maximum magnitude value in whole image m_{max} (f is usually between 0.0001 and 0.01):

$$t_i = i * f * m_{max} \quad (1)$$

Variable value n_i is equal to the number of neighboring pixel magnitudes which differ from central pixel magnitude less than threshold value:

$$n_i = |\{m_j : |m_j| <= t_i, j = 1, \dots, 8\}| \quad (2)$$

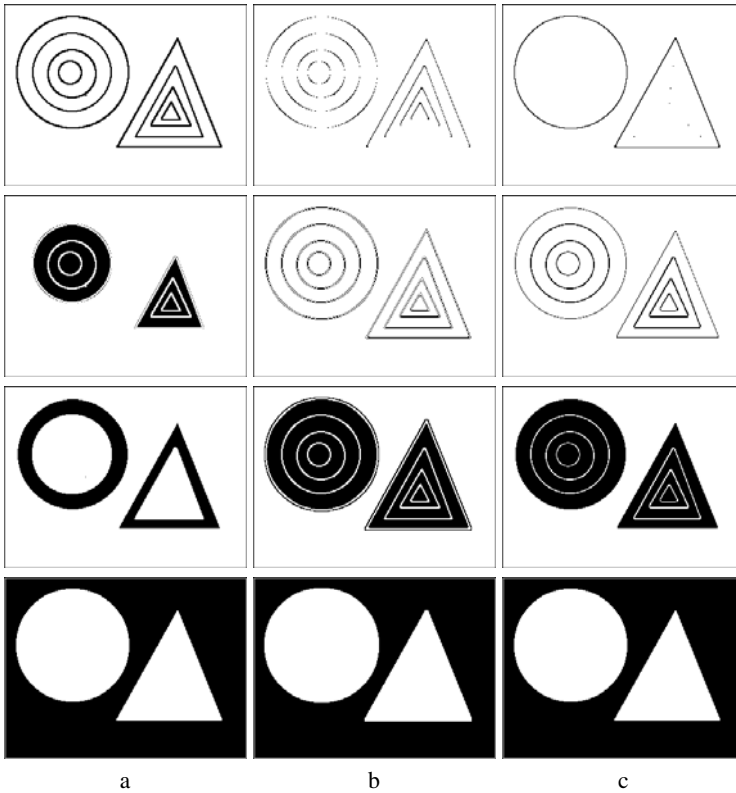


Fig. 2 Synthetic image processed by GCCA supported decomposition: a) normalized gray variables version, b) magnitude variables version, c) normalized magnitude variables version; cluster number is predefined equal to 4

The family of thresholds is linear dependent in this case. Thresholds can be derived from non-linear dependencies as in [9].

In this paper reduction of variable complication is presented. Family of thresholds is unnecessary now and process of establishing of the factor f value is avoided. Instead, a sequence of image gray levels $v_j, j = 1, \dots, 8$ in 8-neighborhood is used. This sequence is ordered according to nondecreasing values:

$$v_1, \dots, v_8 : v_j \leq v_{j+1}, j = 1, \dots, 7 \quad (3)$$

Ordered sequence $v_j, j = 1, \dots, 8$ can represent magnitude values in the neighboring pixels as well. Left column of figure 2 presents subimages obtained with the aid of normalized gray level values, i.e. each variable value of neighboring pixel is derived from gray value of the considered pixel and so ten values of variables used are as follows: $g, m, g_j/g, j = 1, \dots, 8$. The data table has 74576 objects due to magnitude



Fig. 3 Images: a) crossroad, b) Lena , c) NMR fragment

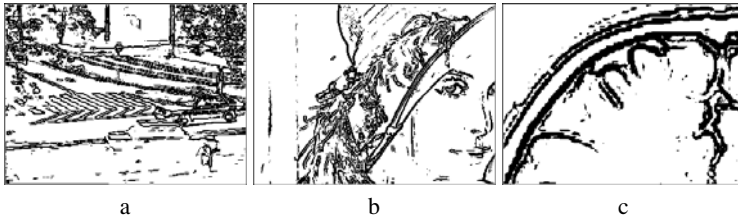


Fig. 4 Gradient magnitude thresholding applied to images from Figure 3

values are not obtainable for pixels adjacent to image boundary. For the synthetic image optimal number of clusters is 4 which results for example from overrepresentation map (not shown for the lack of place). Pixels in subimages are located in the reordered data table near each other due to their similarity of whole description in the transformed table. The clustering procedure determined subsets with alike objects (pixels) so structures in the image can be distinguished. In the synthetic image the clustering procedure separates regions with the same gray value. Also boundary pixels are gathered together. In addition small dark point near the second circle boundary at -45 degrees has its own small representation.

The second and third column of subimages in Figure 2 are obtained using following sets of variable values: $g, m, m_j, j = 1, \dots, 8$ and $g/255, m, m_j/m, j = 1, \dots, 8$. In the first subimage in the middle column directional sensibility can be seen. Pixels with gradient magnitude direction 0 and 90 degree are shifted to second cluster. These pixels have S_x or S_y component of Sobel operator only. Pixels occurs more similar to those from second cluster. Such exceptions are rare in natural images (approximately one hundred in whole image) which is less than 0.13 percent of image pixel amount. The influence of such pixels on image decomposition (e.g. incorrect clustering) can be neglected. Decomposition visualized in third column in Figure 2 involves normalization, magnitudes of neighboring pixels are divided by the magnitude of considered pixel. Directional sensibility disappears due to pixels belonging to contour at any direction become more similar to each other in their record description.

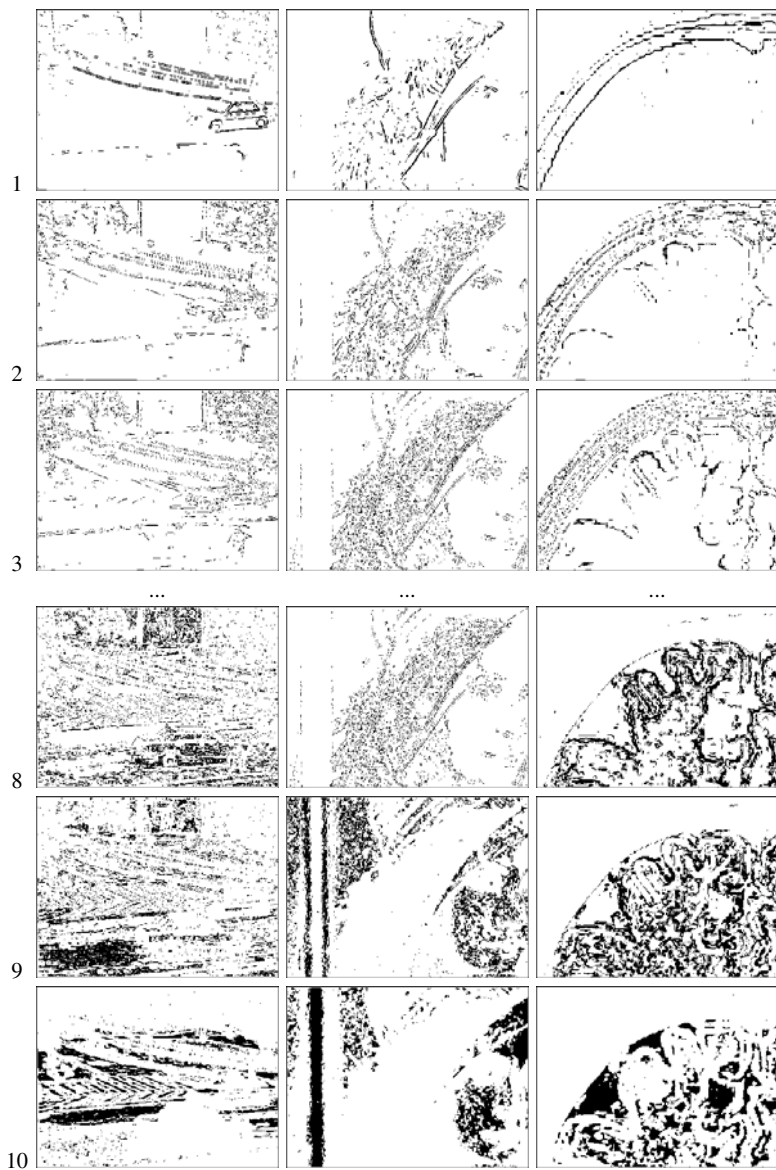


Fig. 5 Structures revealed in images from Figure 3 using decomposition supported by GCCA; at the top of each column first three subimages (out of ten), at the bottom last three subimages

3 Examples of Image Decomposition

In Figure 3 are three test images: crossroad image, Lena (fragment of known Lena image) and fragment of NMR image. All images have space resolution 320×240 and depth resolution 255 like the synthetic image from Figure 1.

Gradient magnitude thresholding is a classic edge detection basis. Figure 4 presents magnitude thresholding for test images performed for threshold value 15. Figure 5 presents GCCA decomposition: crossroad image is processed with normalized gray level values $g, m, g_j/g, j = 1, \dots, 8$, Lena image with magnitude values $g, m, m_j, j = 1, \dots, 8$ and NMR image with normalized magnitude values $g/255, m, m_j/m, j = 1, \dots, 8$.

Table 1 Variable values ordering after GCA

crossroad	m	m_8	m_7	m_6	m_5	m_4	m_3	g	m_2	m_1
Lena	m_8	m_7	m_6	m_5	m	m_4	m_3	m_2	m_1	g
NMR	m_8	m_7	m_6	m_5	m	m_4	m_3	m_2	m_1	g

Table 1 shows variable values ordering after GCA for each test image. The ordered list together with overrepresentation maps allows to make some observations concerning image grade decomposition. Visualization of first three clusters and last three clusters is shown in the Figure 5. First clusters gather pixels with greater overrepresentation for variable values m, n_8, n_7 for crossroad/normalized gray level version, n_8, n_7, m_6 for Lena/magnitudes version and NMR/normalized magnitude version. Values n_8, n_7, m_6 contain the greatest values of v_j . First clusters show in turn underrepresentation for variable values n_2, n_1 . It may also mean greater differences among n_i than expected. This involves that edges may appear in such locations. Subimages at the bottom of Figure 5 contain pixels with opposite tendency: underrepresentation occurs for variables m and n_8, n_7 and overrepresentation for variables n_2, n_1 for crossroad, for variables n_2, n_1, g for Lena and NMR. In these clusters pixels form large regions. Bottom subimages can be compared at some extent to result of split and merge method shown in Figure 6 as well as top subimages to result of gradient thresholding shown in Figure 4.



Fig. 6 Split and merge algorithm applied to images from Figure 3

The number of predefined clusters is chosen arbitrarily. Ten is great enough at a moment but not too big to easily manage subimages.

4 Conclusions

In developed task complication of variables decreases when the thresholds family with the parameter has been withdrawn. Parameter selection is troublesome and its elimination is always advantageous. Described variable value variants come directly from the neighborhood of the pixel. Intuitive interpretation is of great value too. Normalization allows consider similarity of the pixel and its neighborhood more precisely adapting to a change of intensity diversity. Recognition of similarity and differentiation of objects in data table is just the task for which GradeStat is destined. Described variable values collaborate with GradeStat.

The number of pixel clusters is fixed. This number does not depend on number of existing uniform regions in the image. Adjacent regions in visualizing subimages could be better separated when the number of clusters/subimages increases flexible. Special version of GradeStat dedicated to image processing should solve problem and make possible to connect next steps of processing.

To estimate quality of decomposition a comparison could be used to images of Berkeley Segmentation Dataset [12], a set of 300 images and complementary set of boundary images produced by humans. Authors presents a Local and a Global Consistency Measures based on sums of numbers of pixels belonging to corresponding regions in one segmentation and excluded in another. Other applied measure is graph precision vs. recall [2] which are the probability that a machine-generated boundary pixel is a true boundary pixel and the probability that a true boundary pixel is detected. It is used to performance evaluation of four different algorithms. Both measures are intend to check after merging obtained homogenous regions into bigger ones with some of algorithms described in [1] [4] [4] [16].

References

- [1] Estrada, F.J., Jepson, A.D., Chennubhotla, C.: Spectral embedding and min-cut for image segmentation. In: British Machine Vision Conference (BMVC 2004), London, pp. 317–326 (2004)
- [2] Estrada, F.J., Jepson, A.D.: Benchmarking image segmentation algorithms. *International Journal of Computer Vision* 85(2), 167–181 (2009)
- [3] Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Trans. PAMI* 24(5), 603–619 (2002)
- [4] Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. *International Journal of Computer Vision* 59(2), 167–181 (2004)
- [5] Grabowska, G., Wiech, M.: Grade analysis of data from the European Economic Survey 2005 on Economic Climate in Polish Servicing Sector. *Control and Cybernetics* 38, 783–810 (2009)
- [6] GradeStat web site, <http://gradestat.ipipan.waw.pl/>

- [7] Grzegorek, M.: Image Decomposition by Grade Analysis – an Illustration. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.) *Computer Recognition Systems*. ASC, Springer, Heidelberg (2005)
- [8] Grzegorek, M.: Homogeneity of pixel's neighborhoods in gray level images investigated by the Grade Correspondence Analysis. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.) *Computer Recognition Systems*. ASC. Springer, Heidelberg (2007)
- [9] Grzegorek, M.: Variables Applied in a NMR Image Decomposition with the Aid of GCCA. *J. of Medical Informatics & Technologies* 12, 183–187 (2008)
- [10] Grzegorek, M.: Grade Differentiation Measure of Images. In: Kurzynski, M., Wozniak, M. (eds.) *Computer Recognition Systems 3*. AISC, vol. 57, pp. 63–70. Springer, Heidelberg (2009)
- [11] Kowalczyk, T., Pleszczyńska, E., Ruland, F.: *Grade Models and Methods for Data Analysis, With Applications for the Analysis of Data Populations*. Studies in Fuzziness and Soft Computing. Springer, Heidelberg (2004)
- [12] Martin, D., Fowlkes, C., Tal, D., Malik, J.: A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In: *IEEE International Conference on Computer Vision (ICCV 2001)*, Vancouver, Canada, pp. 416–425 (2001)
- [13] Pleszczyńska, E., Wiech, M.: Analysis of medical and demographical data aided by programme GradeStat. Institute of Computer Science PAS, The Children's Memorial Health Institute (2005) (in Polish)
- [14] Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. PAMI* 8(22), 731–737 (2000)
- [15] Szeliski, R.: *Computer Vision Algorithms and Applications*. Springer, Heidelberg (2011)
- [16] Wiech, M., Szczesny, W.: GradeStat - noncommercial statistical application using grade algorithms and methods to make synthesis of information. In: Karwowski, W., Orłowski, A. (eds.) *Information Systems in Management III*. Warsaw University of Life Sciences Press (2009)

Analysis of Inter-rater Agreement among Human Observers Who Judge Image Similarity

Krzysztof Michalak, Bartłomiej Dzieńkowski, Elżbieta Hudyma,
and Michał Stanek

Abstract. In this paper a problem of inter-rater agreement is discussed in the case of human observers who judge how similar pairs of images are. In such a case significant differences in judgment appear among the group of people. We have observed that for some pairs of images all values of similarity ratings are assigned by various people with approximately the same probability. To investigate this phenomenon in a more thorough manner we performed experiments in which inter-rater coefficients were used to measure the level of agreement for each given pair of images and for each pair of human judges. The results obtained in the experiments suggest that the variation of the level of agreement is considerable among pairs of images as well as among pairs of people. We suggest that this effect should be taken into account in design of computer systems using image similarity as a criterion.

1 Introduction

Due to significant growth of the volume of image resources stored in various computer systems there is an increase in the need to search through image collections and to find images that fit user's needs [5]. One of the possible approaches is to make the computer system look for images the most "similar" or the most "dissimilar" to what users select as examples [13]. Unfortunately, there is no easy way to define what "similar images" means, because many criteria (such as the main object, colour, picture style, etc.) can be taken into account. This rises a question: how do

Krzysztof Michalak · Bartłomiej Dzieńkowski · Elżbieta Hudyma · Michał Stanek
Wrocław University of Technology,
Institute of Informatics
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
e-mail: krzysztof_michalak@poczta.onet.pl,
{Bartlomiej.Dzienkowski, Elzbieta.Hudyma,
Michal.Stanek}@pwr.wroc.pl

various people react when asked to judge similarity of the same pair of images? To find an answer to this question we have performed a study in which people judged the level of similarity in a series of image pairs. Data gathered in this study were used to investigate how consistent are the scores assigned by human judges.

2 Experiment Design

User-assigned image similarity scores can be useful as a reference for comparing scores given by automated image similarity assessment methods and also can be used to answer the question: how consistent people as a group are in judging image similarity. Both tasks require a statistical sample containing numerical scores for pairs of images.

In order to produce a consistent result set the experiment was performed on predefined image sets from which all possible pairs were generated. Pairs were only generated from images within the same image set. No cross-set scoring was performed. In the experiments described in this paper 5 image sets were used:

- **city** – images of urban landscapes, buildings and building details.
- **impression** – abstract art images.
- **meadow** – images of meadows in various surroundings (city, mountains, etc.).
- **nature** – natural landscapes, plants and animals.
- **zoo** – images of animals (all of them mammals).

Images in each set were obtained by entering the corresponding keyword ("city", "impression", etc.) into the search box on the Flickr website [2]. Each of the image sets contained the first 30 images returned in the result set, from which all 435 pairs were generated. Examples of images from each image set are presented in Figure 1. All images can be viewed at [1].



Fig. 1 Examples of images used in the experiments

The group of people to whom the image pairs were presented was selected from the University employees and students. People from this group have similar background in aspects of education, cultural profile etc. Each image pair was rated by 30 different people who were **not** instructed on what criteria to use when assigning scores. The scoring process was performed using the *Similaris* web tool, further described in Appendix A. Using this tool image pairs were consecutively

presented to the user. For each image pair a 5-grade scale was provided on which users were asked to mark their votes. The scale extended from "not similar" to "very similar" with more neutral grades in between. In computations these grades were converted to numerical values: 0—not similar, 0.25—slightly similar, 0.5—average similar, 0.75—similar, 1—very similar.

3 Results

Already at the first glance at the score charts we noticed that the distribution of grades varies significantly from one image pair to another. There are image pairs for which all votes are almost identical and such pairs for which the votes distribution is uniform. Identical votes occur more often for image pairs containing totally dissimilar images than for other cases, because it is relatively easier to agree that two images are not at all similar than to agree on the same grade when there is some similarity.

Image pairs visible in Figure 2 present the three most characteristic cases of grades distribution. The first case is an image pair from the "nature" image set for which all judges said the images are not similar at all. The second case is an image pair from the "city" image set for which equally many people have given each of the grades. The third case is an image pair from the "zoo" image set for which people voted mostly for "very similar" and some of them for "similar".

Nature image set
Grades distribution
[30, 0, 0, 0, 0]



City image set
Grades distribution
[6, 6, 6, 6, 6]



Zoo image set
Grades distribution
[0, 0, 0, 4, 26]



Fig. 2 Image pairs with characteristic distributions of the grades

To further investigate the matter of varying judgment of image similarity among users we calculated several coefficients that quantitatively summarize inter-rater agreement [4] in a series of observations. Inter-rater agreement coefficients such as Cohen's kappa K [6, 16], weighted kappa K_w [3, 7] and Fleiss' kappa K_F [9] are used in psychology - to measure agreement among many raters [9], in medicine - for measuring the level of agreement when several independent diagnoses are made [12], in computer science - for measuring the level of agreement between text annotators [15] and also in music information retrieval for comparing various methods of song similarity assessment [8]. As these measures have been used by other authors to numerically express the level of agreement among various raters and annotators (human as well as automatic) they seem very well suited for analyzing the behaviour of human raters as we wanted to do in our research. If the level of agreement between two raters is measured, the commonly used statistical measures of distribution spread such as variance or standard deviation do not give meaningful results because they are equally sensitive to differences between scores assigned by the same judge to different objects (which reflect the diversity of the data set) as they are to differences between scores assigned by different judges to the same object. Inter-rater agreement measures used in this case are designed so that they focus on measuring only the differences between the judges. In the literature well established guidelines for interpretation of values of inter-rater agreement measures can be found (see Tables 3 and 4), and thus the meaning of numerical values can easily be understood.

The inter-rater agreement can be assessed in two ways: we can measure agreement among several judges rating the same image pair and we can measure agreement between any two given judges for all image pairs in a given image set.

The first approach requires a measure of inter-rater agreement that can be calculated for any number of people who assign scores to the same image pair. For this task we used the Fleiss' kappa coefficient [9] which is calculated from score distributions for any number of objects as shown in [8] using equations (1)-(2). The advantage of this coefficient is in our case the fact that it can be calculated in particular for one object (one image pair). The summary of values of the Fleiss' kappa obtained for individual pairs of images in each of the image sets is given in Table 1

$$K_F = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}, \quad (1)$$

where:

$$\bar{P} = \frac{1}{Nn(n-1)} \sum_{i=1}^N \sum_{j=1}^k n_{ij}(n_{ij} - 1), \quad \bar{P}_e = \sum_{j=1}^k \left(\frac{1}{Nn} \sum_{i=1}^N n_{ij} \right)^2, \quad (2)$$

N - number of pairs, n - number of judges, k - number of grades.

The lowest value of Fleiss' kappa coefficient in the "city" image set was recorded for the same image pair for which the most uniform distribution of grades, namely [6, 6, 6, 6, 6], was obtained. This pair of images is presented in Figure 2

Table 1 Values of the Fleiss’ kappa obtained for individual pairs of images

Image set	Mean	SD	min	max
city	0.4885	0.2267	0.1667	1.0000
impression	0.4770	0.2339	0.1645	1.0000
meadow	0.6138	0.2543	0.1703	1.0000
nature	0.5222	0.2489	0.1714	1.0000
zoo	0.5930	0.2609	0.1830	1.0000

As can be seen from the values of standard deviation (SD) in Table 1 the values of Fleiss’ kappa vary greatly in each of the image sets. This is visualized in Figure 3. From the histograms it can easily be seen that the inter-rater agreement changes significantly from one image pair to another.

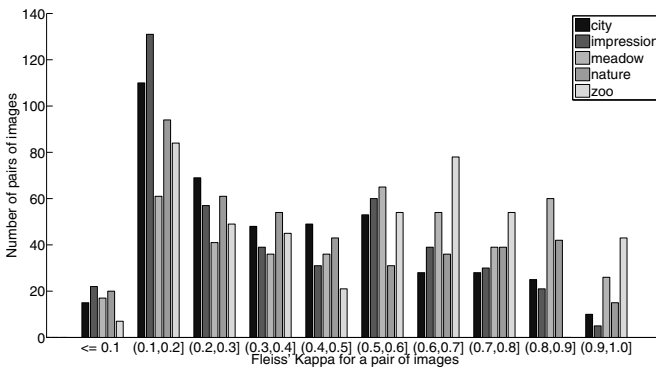


Fig. 3 Histogram of Fleiss’ kappa values

The second method of inter-rater agreement assessment is to measure the agreement separately for each pair of users using scores assigned by these two users to all image pairs in a given image set. In this case there are several coefficients that can be used. The most straightforward is a simple agreement frequency F (the number of times both users have assigned identical scores divided by the total number of scores given by each user). This coefficient, however, is often criticized because a non-zero frequency of agreements is expected to occur purely by chance, so the lower end of the scale is not meaningful.

The Cohen’s kappa coefficient K [6, 16] is aimed at correcting this shortcoming. As shown in [17] it can be calculated using equation (3). Compared to the simple agreement frequency F , values of the Cohen’s kappa coefficient K are scaled from the range $[P_c, 1]$ to $[0, 1]$. If the observed frequency is lower than "pure chance" agreement frequency the value of Cohen’s kappa coefficient may be $K < 0$.

$$K = \frac{P_o - P_c}{1 - P_c} \quad (3)$$

where:

P_o – observed frequency of the same score being given by both users.

P_c – the probability of two scores being equal by "pure chance", calculated using equation (4).

$$P_c = \sum_{i=1}^k P_1(G_i)P_2(G_i) , \tag{4}$$

where:

k – the number of grades (values used for rating),

$P_j(G_i)$ – the probability that the j -th user assigns i -th grade (estimated from sample as a frequency of i -th value among all scores assigned by j -th user).

The weighted kappa coefficient K_w [3, 7] is calculated with higher weights assigned to disagreements in which there is a greater discrepancy between the grades. This measure is therefore better-suited for non-binary ratings such as the one we have in our tests. The weighted kappa coefficient is calculated as shown in [1] using equation (5).

$$K_w = 1 - \frac{\sum w_{ij}P_{oij}}{\sum w_{ij}P_{cij}} , \tag{5}$$

where:

w_{ij} – weight of the disagreement between grades i and j ,

P_{oij} – observed frequency of grade i being assigned by the first user and grade j being assigned by the second user,

$P_{cij} = P_1(G_i)P_2(G_j)$ – "pure chance" probability of grade i being assigned by the first user and grade j being assigned by the second user.

Table 2 contains mean values of inter-rater agreement coefficients obtained for data from the experiments.

Table 2 Mean values of inter-rater agreement coefficients obtained for pairs of judges

Image set	Cohen's kappa	Weighted kappa	Agreement frequency
city	0.1925	0.3292	0.4955
impression	0.1942	0.3055	0.4771
meadow	0.2098	0.3215	0.6138
nature	0.1663	0.2701	0.5222
zoo	0.2739	0.4351	0.5930

The mean values of all coefficients obtained in the experiments are quite low which leads to the conclusion that on average people may have quite different opinions on whether two images are similar or not. By plotting distributions of inter-rater agreement coefficients it can be shown that there is a varying level of agreement

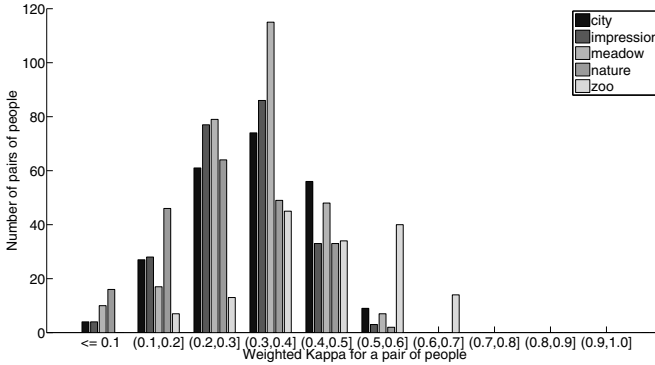


Fig. 4 Histogram of weighted kappa values

in pairs of people. Figure 4 shows distributions of the weighted kappa coefficient for pairs of people in each of the image sets.

From the histograms which present distributions of values of weighted kappa coefficient calculated for pairs of judges it is clear that the level of agreement of image similarity scores varies greatly in the group of people.

Numerical values obtained in the experiments can be interpreted using kappa interpretation guidelines available in the literature. In Tables 3 and 4 interpretations suggested by Fleiss et al. for the Fleiss' kappa coefficient [10] and by Landis and Koch for kappa coefficient [14] are presented. These tables also contain percentages of kappa values obtained for each of the image sets.

Results presented in Tables 3 and 4 show, that the level of agreement among users who judge image similarity is rather low. Fleiss' kappa coefficient can mostly be interpreted as "poor" or "moderate" agreement. The situation is even worse when values of weighted kappa are interpreted. For four of the five image sets values of the weighted kappa coefficient indicate nothing more than "moderate" agreement at best. Only for the "zoo" image set and for a few pairs of people the level of agreement could be interpreted as "substantial".

Table 3 Interpretation of the Fleiss' kappa coefficient according to guidelines presented in [10]. The table also presents the percentage of image pairs for which agreement level is in a given range.

Fleiss' kappa κ_F	Interpretation (agreement level)	city	impression	meadow	nature	zoo
< 0.40	poor	45%	48%	27%	40%	32%
0.40 – 0.74	moderate	41%	37%	33%	38%	32%
0.75 – 1.00	good	14%	15%	40%	22%	36%

Table 4 Interpretation of the weighted kappa coefficient according to guidelines presented in [14]. The table also presents the percentage of pairs of people for which agreement level is in a given range.

Kappa κ	Interpretation (agreement level)	city	impression	meadow	nature	zoo
< 0.00	poor	0%	0%	0%	0%	0%
0.00 – 0.20	slight	15%	16%	12%	31%	5%
0.21 – 0.40	fair	60%	71%	71%	55%	40%
0.41 – 0.60	moderate	25%	13%	17%	13%	49%
0.61 – 0.80	substantial	0%	0%	0%	0%	6%
0.81 – 1.00	almost perfect	0%	0%	0%	0%	0%

4 Conclusion

The experiments described in this paper were aimed at investigating how do people perceive similarity of images. From even the most cursory glance on test results it became obvious that there are significant differences in how various people judge similarity of images.

In order to measure the level of agreement among human judges we used inter-rater agreement coefficients: agreement frequency, Cohen's kappa, weighted kappa and Fleiss' kappa. Using these coefficients we measured the agreement among all users who assigned scores to each given image pair and the agreement between each two judges calculated using scores assigned to all images in each image set.

From the results it can be seen that there exist image pairs that are given similar grades by all users (especially pairs with clearly dissimilar images). However, there are also image pairs for which the scores given by users vary significantly. The most prominent example we have encountered was an image pair presented in Figure 2 for which each of the grades was assigned by exactly the same number of users. Similarly, when the level of agreement is measured in pairs of judges it can be seen that there exist pairs of people who agree to various extent in their scoring of a given image pair set.

From the experiments we can draw a conclusion that the criterion of "image similarity" is not as well defined as it may seem. Thus, in all applications where a computer system is required to produce results based on image similarity it must be taken into account that people may perceive image similarity in very different ways.

The significant differences in the assessment of image similarity suggest that people may evaluate image similarity based on various criteria, such as the main object, image composition, dominating color, image type (portrait, wide angle shot, etc.). It might be worth considering that in real-life applications of computer systems this diversity should be taken into account. In the case of image-search system this can be achieved for example by asking users to explicitly choose what criterion of image similarity should be used (for example: "find images similar **in color** to the one I selected").

Acknowledgment. This work is financed from the Ministry of Science and Higher Education Republic of Poland resources in 2008-2010 years as a Poland-Singapore joint research project 65/N-SINGAPORE/2007/0.

Appendix A - The *Similaris* Tool

The idea of *Similaris* was to create a convenient tool which might help us find an answer to the crucial question: How do people perceive similarity of two pictures?

For each given pair from the test image set the similarity level indicated by the user is stored and then the entire collection of votes can be analyzed. The *Similaris* tool provides visualization of the results, basic statistical indicators such as mean, variance, etc. and allows to export raw data to a CSV file for further analysis. We want to use the answers gathered in such a way to find out what is taken into account when pictures are compared by people and how to adjust image similarity assessment algorithms to better match human behaviour.



Fig. 5 User interface of the *Similaris* tool

Using the *Similaris* tool users can choose one of many picture collections (experiments) and start the test. Picture pairs appear one by one on the screen and the user is asked to select the level of similarity between these two pictures during a predefined period of time. It is possible to submit only one answer for one pair of pictures. A test is finished when all the pairs generated from a given image set were scored by the user.

References

- [1] All images used in the experiments at the AI research group website (2010), <http://ai.ii.pwr.wroc.pl/similaris/pic/PWR-SimilarImages/>
- [2] Flickr - photo sharing (2010), <http://www.flickr.com/>

- [3] Berry, K., Johnston, J., Mielke Jr., P.: Weighted kappa for multiple raters. *Percept Mot Skills* 107(3), 837–848 (2008)
- [4] Bland, J.M., Altman, D.: Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 327(8476), 307–310 (1986)
- [5] Cheng, S.C., Wu, T.L.: Fast indexing method for image retrieval using k nearest neighbors searches by principal axis analysis. *Journal of Visual Communication and Image Representation* 17(1), 42–56 (2006)
- [6] Cohen, J.: A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1), 37–46 (1960)
- [7] Cohen, J.: Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin* 70(4), 213–220 (1968)
- [8] Downie, J.S., Ehmann, A.F., Bay, M., Jones, M.C.: The music information retrieval evaluation exchange: Some observations and insights. In: Raś, Z.W., Wieczorkowska, A.A. (eds.) *Advances in Music Information Retrieval. SCI*, vol. 274, pp. 93–115. Springer, Heidelberg (2010)
- [9] Fleiss, J.L.: Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5), 378–382 (1971)
- [10] Fleiss, J.L., Levin, B., Paik, M.: *Statistical Methods for Raters and Proportions*, 3rd edn. Wiley and Sons, Chichester (2003)
- [11] Geertzen, J., Bunt, H.: Measuring annotator agreement in a complex hierarchical dialogue act annotation scheme. In: *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, Association for Computational Linguistics, Sydney, Australia, pp. 126–133 (2006)
- [12] Jakobsen, K.D., Frederiksen, J.N., Hansen, T., Jansson, L.B., Parnas, J., Werge, T.: Reliability of clinical ICD-10 schizophrenia diagnoses. *Nordic Journal of Psychiatry* 59(3), 209–212 (2005)
- [13] Kherfi, M.L., Ziou, D., Bernardi, A.: Combining positive and negative examples in relevance feedback for content-based image retrieval. *Journal of Visual Communication and Image Representation* 14(4), 428–457 (2003)
- [14] Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. *Biometrics* 33(1), 159–174 (1977)
- [15] Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R., Araki, K.: CAO: A fully automatic emoticon analysis system. In: Fox, M., Poole, D. (eds.) *AAAI*. AAAI Press, Menlo Park (2010)
- [16] Viera, A.J., Garrett, J.M.: Understanding interobserver agreement: the kappa statistic. *Family Medicine* 37(5), 360–363 (2005)
- [17] Yoon, T., Chavarrá, R., Cole, J., Hasegawa-johnson, M.: Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. In: *Proc. ICSLP* (2004)

A Tool for Implementing and Evaluating Image Processing Algorithms inside FPGA Circuits

E. Guzmán, I. García, and M. Mendoza

Abstract. This paper aims to show the design and implementation of an integral environment for developing algorithms for digital images processing and analysis on reconfigurable logic. Our tool was divided in four modules. The first module designs a specific application architecture using a hardware description language. Another module uses the designed architecture for designing a hardware system using the SPIES development method. A third module develops a GUI to interact with the embedded system and show the obtained results for processing images with an algorithm inside a FPGA. A last module enables the user to implement and integrate images processing/analysis algorithms through the developed tool. We believe that our tool can be used for academic and research interests with a high level of success.

1 Introduction

The digital image processing and analysis is an area that allows one to manipulate and to analyze the information contained in a digital image by means of a processor. The digital image processing is a set of methods that are applied to digital images with the goal to improve its quality, and the image analysis includes those methods which the principal goal is to facilitate the search and interpretation of information contained in the image [9]. The scientific and academic interest for image processing and analysis on autonomous systems for solving problems associated to the artificial vision, such as objects recognition, trajectory planning of robots, etc., has grown in the last years. On the other hand, the reconfigurable logic has attractive features to implement applications of artificial vision on embedded systems. Some of these features are: (1) to allow the user to handle a high level of abstraction in the system design; (2) to facilitate the implementation of System on Chip (SoC); (3) to involve a

E. Guzmán · I. García · M. Mendoza
Universidad Tecnológica de la Mixteca,
Carretera a Acatlima Km. 2.5, Huajuapán de León, Oaxaca, México
e-mail: eguzman@mixteco.utm.mx

design paradigm based on software methodologies; and (4) to allow the modeling of parallel architectures that generate high processing speeds; resources optimization; when standardized HDL's is used, portable designs are generated, etc.

Bearing in mind the importance of the role that redeems the area of digital image processing and analysis in our days, and the absence of tools that facilitate the design of new algorithms on reconfigurable logic, this paper presents the design and implementation of a digital system of image processing and analysis based on Field Programmable Gate Array (FPGA) technology. The remaining sections of this paper are organized as follows. The next section alludes to work related to the proposals in this paper. Section 3 describes the design and implementation of the hardware-software tool for digital image processing and analysis within a FPGA; also, in this section an integration method that establishes a sequence of procedures that enables the implementation and performance evaluation of an algorithm over the FPGA is proposed. The experimental results are presented in Section 4; the modeling of the Wavelet Discrete Transform on reconfigurable logic is evaluated. Finally, the Section 5 contains the conclusions of this paper.

2 Related Work

For real-time image processing systems, which need high performance requirement because of a large amount of image data processed. So, special hardware or software technology of parallel processing is particularly important, and FPGA chips have natural advantages for real-time image processing systems because of their specific units on the logical structure. Some related works that have relation with the processing process improvement using FPGA are summarized in Table 1.

Table 1 FPGA related work

Reference	Main contribution
[4]	A system which provides a high level image processing oriented programming model that is automatically implemented on an FPGA-based machine.
[6]	An educational environment (IPXS tools) for design and easy checking of FPGA programmed applications.
[14]	An architecture based on FPGA for embedded applications.
[3]	The MApFPGAs (Multiple Applications for FPGAs) control on the FPGA for image processing in PC.
[5]	An image processing platform based on FPGA. The proposed system uses a Virtex-4 SX35 FPGA device.
[7]	An architecture of a hardware remote laboratory based on using FPGA development boards for digital electronics circuit design.
[12]	A specific tool for the evaluation of applications on FPGAs that includes support of hardware and software.
[11]	A general-purpose, multi-task, and reconfigurable platform for video and image processing using a FPAG parallel processing.

Our proposal is a tool that offers a solid base in the design and implementation of hardware prototypes that include algorithms of processing and/or analysis of images. The principal aim of this tool is to speed up the design and modeling of an algorithm and to allow visualizing the results before implementing on the final hardware system. The principal fields of operation of our tool include ambiences, academicians and investigation, related to the areas where the digital image processing and the logic reconfigurable converge. The tool design was divided in four modules:

Module 1. In the first module, the design and modeling of specific application architecture, using a Hardware Description Language (HDL), focused on the digital image processing was realized.

Module 2. In the following module, based on the designed architecture and using the SPIES method [8], the design and implementation of an embedded system was performed.

Module 3. A third module develops a GUI to interact with the embedded system and show the obtained results for processing images with an algorithm inside a FPGA; the FPGA tool' interface was developed using the VY method proposed in [15]. The GUI implementation was realized on the Integrated Development Environment C++; the USB standard is used as communications protocol with the embedded system.

Module 4. Finally, the integration method module on physical layer establishes a sequence of procedures that enables the implementation and performance evaluation of an algorithm over the FPGA; these procedures involve both to the specific application architecture of the first module, and to the GUI developed in the module three.

3 The Tool Architecture

Fig. 1 shows that the basic architecture of FPGA tool is divided into two layers and it is composed by four modules:

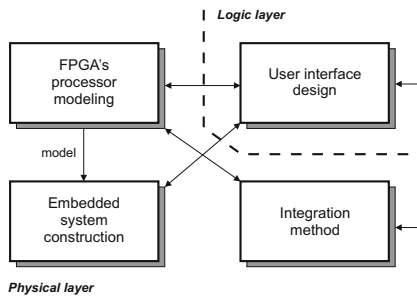


Fig. 1 Modules of FPGA tool

The *physical layer* establishes the configuration of embedded system used to process, analyze, and retrieve information stored in a memory system. The layer is composed by the following modules:

- *The FPGA's specific application architecture modeling*: This module follows a descendent methodology for designing and modeling a basic structure of specific application architecture, called Hardware Applications Manager (HAM), see section 3.1. This architecture is integrated with components for interfacing to User Interface, connecting with the system memory and supervises the execution of image processing algorithms. This module uses the *integration method module* for adding new algorithms to process and/or analyze images.
- *Embedded systems construction*: Using the SPIES method, an embedded system is constructed using the FPGA's HAM modeling. The HAM modeling and the embedded system are designed to allow to add new peripheral to the system and to facilitate the integration of new algorithms to the architecture.
- *The Integration method*: This module provides the capability for integrating new algorithms for processing and/or analyzing to HAM architecture and enables its evaluation. The module incorporates an integration method that establishes the activities sequence to specify, integrate and evaluate the performance of a new algorithm (see Section 3.5 for a detailed explanation).

The *logical layer* allows to select the image that will be processed, to send it to embedded system and enables visualization of images processed inside the images processor. The layer is composed by the following module:

- *User interface design*: All the processing work inside the FPGA is showed by the user interface separating the algorithms parameters from processed images.

To make possible the implementation of the image processing tool it was used a FPGA Spartan 3E-500 from Xilinx Company. This FPGA is used to store the image processing algorithms modeled by user. Fig. 2 shows the block diagram of the proposed tool.

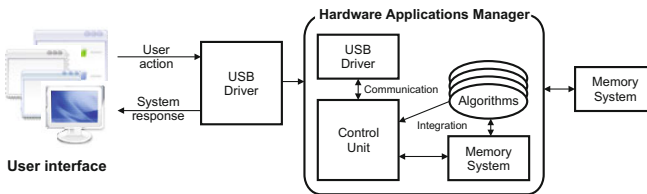


Fig. 2 Architecture of FPGA Tool

3.1 Hardware Applications Manager

This module belongs to the physical layer and contemplates the design and implementation of specific application architecture on logic reconfigurable, called HAM. HAM architecture is integrated by four subsystems:

- *Control unit*. This subsystem administers the system resources and the functions developed by the HAM architecture; e.g., it verifies that the communication between the system and the user interface is correct, the image reception to process, the results transmission to the user interface, etc.

- *Memory System driver.* This subsystem controls the access to the memory system when the HAM architecture or a user's algorithm request a write or read information.
- *USB driver.* This subsystem controls the information transfer between the HAM architecture and the user interface across an USB external device.
- *Algorithms.* The HAM architecture has integrated some images processing and analysis algorithms; furthermore, new algorithms can be included; each of these new algorithms must be defined inside the HAM architecture as a new subsystem.

In the HAM architecture implementation, the board Nexys 2 and Xilinx ISE Foundation 8.2i EDA tools was used. The Nexys 2 board was designed by Digital Inc. company and it is based on Xilinx FPGA Spartan 3E-500. The Xilinx ISE Foundation is an integrated software environment used in the digital systems design.

3.2 *Embedded System Development Method*

FPGA tool was developed concurrently with the hardware and other components of the product in the Product Development phase of the SPIES method [8]. Development processes specification belongs to engineering process field of TSP [10] and process areas from Level 2 of CMMI-DEV v1.2 [13]. Engineering process field is usually related to requirement development, technical solutions, etc. Therefore, SPIES provides concrete technical development specification for embedded applications that includes templates and guidelines. In addition, besides software processes, hardware related processes are emphasized here. Using SPIES and the HAM architecture as central element, an embedded system focused to digital image processing and analysis is constructed; it represent the second module of our tool, called *Embedded systems construction*.

3.3 *User Interface*

According to [2], Graphical User Interface (GUI) for applications of embedded devices has been increasingly in demand. Our FPGA tool' interface was developed using the VY method proposed in [15]. The VY method modifies the traditional way of GUI development for embedded applications by adding a simulator. The simulator was designed to simulate the graphic functions and processing algorithms into the HAM architecture; in addition a graphic interface which is implemented by Graphical Device Interface on PC was developed. The logic part of GUI includes focusing, state transaction, and user functions. The simulator was implemented on the FPGA including image processing algorithms. The second important feature of the simulator is the mechanism of the message-driven process. The processes are always waiting for selection of processing algorithms from the user, and when received, the image is stored in memory, which finally gets processed as shown in Section 4.

3.4 Communication Protocol

The communication protocol allows the interaction between the user's interface and the embedded system. Fig. 3 describes the communication protocol used to establish the data interchange and that is transparent for users. A brief explanation of each package can be resumed as follows:

- The *Send image package* is used by user interface to send an image to FPGA. Fig. 3 (A) shows that the package structure is composed by four attributes: ID, CTR, DIR and DATA.
- The *Implement algorithm package' structure* is designed to enable user interface to indicate what algorithm should be applied to image inside the FPGA. According to Fig. 3 (B) this package is composed by three attributes: ID, DIR and CTR.
- The *Results package* enables to FPGA send the results to user interface. This package is designed to facilitate the understanding of obtained result and present it and adequate form. The package is composed by three attributes: ID, CTR and DATA.
- The *Communication package* initializes the USB port and solicits an affirmation to FPGA related to the establishment of the communication. This package can be used at any moment to verify that the communication is established. ID and CTR attributes constitutes the package structure.
- The *Confirmation package* is sent when the FPGA receives a Communication package with the A5 Hex value in the CTR attribute. Fig. 3 (E) shows that the structure package is composed by two attributes: ID and CTR.

3.5 The Integration Method

The integration method module on physical layer establishes a sequence of procedures that enables the implementation and performance evaluation of an algorithm

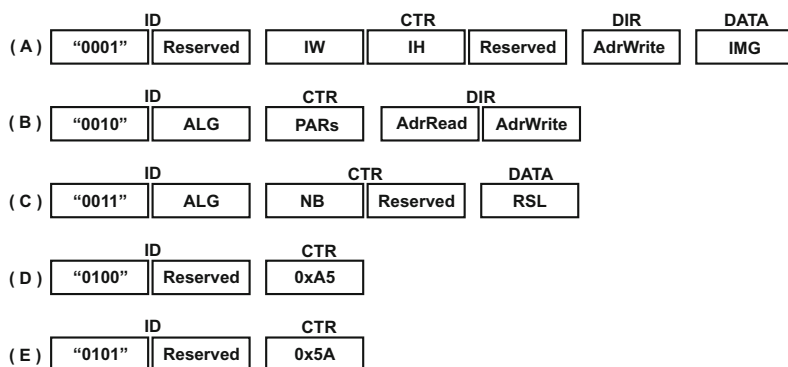


Fig. 3 (A)Send image package (B)Implement algorithm package (C)Results package (D)Communication test package (E)Confirmation package

over the FPGA. With the integration of this module into the system, a tool for design is obtained for pre visualizing the results of use an algorithm before its final application. Thus, the development cycle for image processing software is considerably reduced.

4 Experimental Results

This section shows how to integrate a new algorithm to the proposed tool; also, the results of its evaluation using the user interface are presented. The algorithm that will be integrated to the tool is the Discrete Wavelet Transform (DWT). The DWT is an algorithm widely used in the digital image processing; example of this fact, the DWT represents the base of the JPEG2000 standard [1]. The integration of the DWT algorithm to the proposed tool it is done by means of the integration method exhibited in the Section 3.5.

Phase 1. *Algorithm analysis and specification.* The specifications to consider for the algorithm integration are the following ones:

- Only an algorithm is evaluated, the DWT; it uses filters Haar.
- Grayscale images will be processed.
- The processing result is a new image with the same number of elements as original.
- The memory system space used to store the original image is 0x000000--0x00FFFF and the obtained coefficients will be stored in 0x010000-0x01FFFF.
- Parameters to consider in the DWT algorithm: number of levels.

Phase 2. *Integrated conceptual design.* In this phase, the conceptual design of the DWT algorithm, inside of the HAM architecture, and its corresponding GUI are developed. Also, with the information spilled in the specifications the communications protocol packages are formed. The following steps are performed:

- *Conceptual design of the DWT algorithm.* The inputs to the DWT module are pixels of the image to process and the outputs are the transform coefficients. This module is formed by four internal blocks; every block forms a Haar filter (LL, LH, HL o HH) that generates the output coefficients of each sub-band that integrate the resultant image.
- *Conceptual design of the DWT algorithm user interface.* In this sub-phase, the DWT algorithm interface is adapting to the user's interface described in the Section 3.3.
- *Packages definition: send image package, implement algorithm package and results package.* The send image package specifies the image dimension to process (IW and IH), the image data (IMG) and the space of the memory system where they will be stored (0000000). The implement algorithm package is used to identify to the DWT algorithm (0001); also, this package specifies the address of the image data to process (0000000) and the address where the results must be stored (0010000). The results package includes the data number that integrates the result and the data result.

Phase 3. Parallel implementation. Based on conceptual design, defined in the Phase 2, in this phase the modeling of the DWT algorithm and its adaptation to the user's interface are implemented. The DWT algorithm interface is developed and adapted to the user's interface following the VY method; the DWT algorithm interface is a Single Document Interface (SDI) application contained in the user's interface and accessible across its principal menu.

Phase 4. Component integration. When the user's interface has been modified and adapted for the visualization of the DWT results in a new SDI application and a new version of the HAM architecture, which includes to the DWT algorithm, has been programmed in the FPGA, both components are integrated to evaluate the new version of the Tool to Support Image Processing and Analysis using Reconfigurable Logic.

Phase 5. DWT Algorithm evaluation. Fig. 4 shows the results obtained when our FPGA Tool applied the DWT algorithm on 8 bits/pixel grayscale Lena image with size 256x256.

From the obtained results is possible to observe an image formed by 4 components or sub-bands: LL, HL, LH, HH. The LL sub-band contains low pass information, which is essentially a low resolution version of the image. Sub-band HL will contain low pass information vertically and high pass information horizontally, and sub-band LH will contain low pass information horizontally and high pass information vertically. Finally, sub-band HH will contain high pass information in both directions. These results coincide with the theory; therefore, the incorporation of the algorithm wavelet to our tool has been successful.

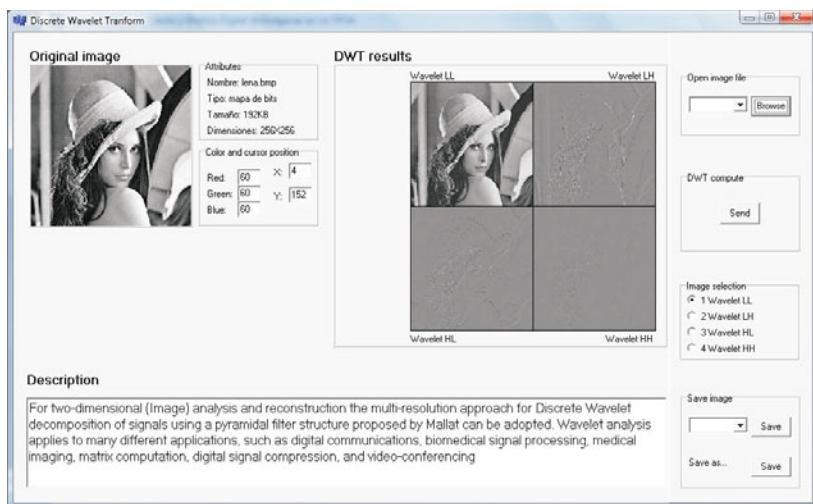


Fig. 4 DWT algorithm interface

5 Conclusions

We are providing a hardware-software tool for digital image processing and analysis within a FPGA. This design would be useful to provide a solid basis for designing and implementing autonomous prototypes that include algorithms for digital image processing and/or analysis. The main goal of our tool is to make more agile the design, modeling and evaluation of an algorithm inside the FPGA and visualize the obtained results before that the implementation occurs. The proposed tool in this paper has been used to evaluate diverse algorithms of images processing and analysis showing a high efficiency.

References

- [1] Acharya, T., Ray, A.K.: *Image Processing: Principles and Applications*. John Wiley and Sons, USA (2005)
- [2] Bai, Y.W., Tsai, F.E.: Design and Implementation of a Table-based GUI for MP3 players. *IEEE Transactions on Consumer Electronics* 53(2), 554–560 (2007)
- [3] Baobre, L., Losada, R., Alvarez, J.: Sistema de Gestin de Aplicaciones Implementadas en FPGAs. Presented in: III Jornadas sobre Computacin Reconfigurable y Aplicaciones (2003) (in Spanish)
- [4] Benkrid, K., Crookes, D., Bouridane, A., Corr, P., Alotaibi, K.A.: A High Level Software Environment for FPGA Based Image Processing. In: *Seventh International Conference on Image Processing and its Applications*, pp. 112–116 (2002)
- [5] Birla, M.K.: FPGA Based Reconfigurable Platform for Complex Image Processing. In: *IEEE International Conference on Electro/information Technology*, vol. EIT 2006, pp. 204–209 (2006)
- [6] Cuenca, S.: *IPXS v. 2.2 User Guide*. Information Technology and Computation Department. University of Alicante, Spain (2000)
- [7] El Medany, W.M.: Remote Laboratory for Hardware E-Learning Courses. In: *International Conference on Computational Technologies in Electrical and Electronics Engineering, SIBIRCON 2008*, pp. 106–109 (2008)
- [8] Garcia, I., Pacheco, C., Herrera, A.: Defining a Software Process Improvement-based Methodology for Embedded Systems Development. In: *Electronics, Robotics and Automotive Mechanics Conference, CERMA 2010*, pp. 120–125 (2010)
- [9] Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd edn. Prentice Hall, Englewood Cliffs (2002)
- [10] Humphrey, W.: *Introduction to Team Software Process*. Addison-Wesley, Reading (2000)
- [11] Li, J., He, H., Man, H., Desai, S.A.: General-Purpose FPGA-Based Reconfigurable Platform for Video and Image Processing. In: Yu, W., He, H., Zhang, N. (eds.) *ISNN 2009*. LNCS, vol. 5553, pp. 299–309. Springer, Heidelberg (2009)
- [12] Ramirez, A., Gadea, R., Colom, R., Daz, J.: A Mixed Hardware/Software SOFM Training System. *Revista Computacin y Sistemas* 11(4), 349–356 (2008) (in Spanish)
- [13] Software Engineering Institute. CMMI for Development (CMMI-DEV, V1.2). CMU/SEI-2006 TR-008, Software Engineering Institute, Carnegie Mellon University (2006)

- [14] Tatas, K., Siozios, K., Vasiliadis, N., Soudris, D.J., Nikolaidis, S., Siskos, S., Thanailakis, A.: FPGA Architecture Design and Toolset for Logic Implementation. In: Chico, J.J., Macii, E. (eds.) PATMOS 2003. LNCS, vol. 2799, pp. 607–616. Springer, Heidelberg (2003)
- [15] Yang, L., Choi, Y., Seo, C., Yang, T., Kim, M.: Design of VY: A Mini Visual IDE for the Development of GUI in Embedded Devices. In: International Conference on Software Engineering Research, Management and Applications, SERA 2007, pp. 625–632 (2007)

Textural Features for Scribble-Based Image Colorization

Michał Kawulok, Jolanta Kawulok, and Bogdan Smolka*

Abstract. In this paper we propose how to exploit image textural features to improve scribble-based image colorization. The existing techniques work by propagating color from the user-added scribbles over the whole image. The color propagation paths are determined so as to minimize the luminance changes integrated along the path. In our method, at first linear discriminant analysis is performed over the scribble pixels to extract discriminative textural features (DTF). Our contribution to image colorization lies in using DTF for the path optimization instead of the luminance. The colorization results presented in the paper explain and confirm the method's robustness compared with the alternative existing techniques.

1 Introduction

Image colorization is an automated process of converting a monochrome image into a color image based on some arbitrary knowledge delivered by a human observer. This initial information may take a form of user-added scribbles which indicate the color of selected image pixels. Based on these few scribbles, the whole image is colorized afterwards. Image colorization is usually used to add colors to old, monochrome photographs or to enhance visual attractiveness of images with faded colors. Its applications include as well marking regions of interest in medical images or colorizing aerial and IR images.

Scribble-based colorization attracts considerable attention and it has been extensively studied. Levin et al. [1] formulated an optimization problem based on an assumption that neighboring pixels of similar intensity should have similar

Michał Kawulok · Jolanta Kawulok · Bogdan Smolka
Faculty of Automatic Control, Electronics and Computer Science,
Silesian University of Technology,
Akademicka 16, 44-100 Gliwice, Poland
e-mail: {mkawulok, jkawulok, bsmolka}@polsl.pl

* This work has been supported by the Polish Ministry of Science and Higher Education under R&D grant no. N N516 374736 from the Science Budget 2009-2011.

color values under the limitation that the colors indicated in the scribbles remain the same. Yatziv and Sapiro [2] proposed a method for determining propagation paths in the image by minimizing geodesic distances from every scribble. Based on the distances from each scribble, pixel color is obtained by blending scribble chrominances. In other works, the propagation is performed using different techniques, for example with probabilistic distance transform [3] or based on random walks [4]. Among other related colorization techniques, it is worth to mention the color transfer method [5]. A grayscale image is colorized based on a given reference color image. This method matches textural and luminance information and can be performed automatically, but gives better results with user assistance.

In the work reported here we have investigated how to exploit the textural information to improve the colorization result. Although the existing techniques work well for colorizing plain areas, they fail for rough, textured regions. Therefore, we transform the image pixels into a discriminant textural feature space, in which the color propagation is performed. This is the main contribution of our work, while in alternative approaches the costs are obtained based on the difference in pixel luminance [2]. The propagation itself, as well as the chrominance blending, is executed following the conventional techniques presented in Section 2. Our method for computing the textural features and color propagation using the obtained feature space is described in Section 3. The obtained colorization results are shown and discussed in Section 4, and the conclusions are presented in Section 5.

2 Color Propagation Paths and Chrominance Blending

In order to colorize a monochromatic image Y based on a set of n initial scribbles $\{S_i\}$, $i = 1, \dots, n$, first it is necessary to determine the propagation paths from each scribble to every pixel in the image. A path from a pixel x to another pixel y is defined as a discrete function $p(t) : [1, l] \rightarrow Z^2$, which maps a position t in the path to the pixel coordinate. The position is an integer ranging from 1 for the path beginning ($p(1) = x$) to l for its end ($p(l) = y$). Also, if $p(i) = a$ and $p(i+1) = b$, then a and b are neighboring pixels.

The propagation paths from a scribble to every pixel are determined by minimizing a *total path cost*:

$$C(p) = \sum_{i=1}^l \rho \{p(i), p(i+1)\}, \quad (1)$$

where ρ is a *local cost* between two neighboring pixels and $l > 1$ is the path length. An image is considered as a graph and the cost minimization is performed using Dijkstra algorithm [6]. The path route depends mainly on how the local costs are computed. Following Yatziv's approach [2], the local cost is obtained by projecting the luminance gradient onto a line tangent to the path direction. This means that the cost is proportional to the difference in luminance between the neighboring pixels.

Chrominance of each pixel is determined based on the propagation paths from every scribble. Its value is computed as a weighted mean of scribbles' colors with the weights obtained as a function of the total path cost. The chrominance is calculated as a weighted mean of scribbles' colors with the weights obtained as a function of the total path cost. Usually two or three strongest components are taken into account which provides a good visual effect of smooth color transitions. The final color value $v(x)$ of a pixel x is obtained as

$$v(x) = \sum_i v_i w_i(x) / \sum_i w_i(x), \quad (2)$$

where v_i is the chrominance of i -th scribble and $w_i(x)$ is its weight in pixel x . We use $YCbCr$ color space and calculate color values separately for C_r and C_b channels. The weights are obtained as

$$w_i(x) = (C_i(x) + 1)^{-2}, \quad (3)$$

where $C_i(x)$ is the total path cost from i -th scribble to pixel x .

3 Textural Features for Image Colorization

Regions of uniform texture quite often have similar chrominance. Following this observation, the texture may be an important source of information in image colorization. Unfortunately, this is neglected by many existing techniques, which assume that the chrominance boundaries are correlated with the luminance changes rather than with the texture. Following this assumption, the raw pixel values in luminance channel are used as the color propagation domain [2, 7]. In our work we focus on finding a better color propagation domain which takes into account the textural features rather than the luminance exclusively.

Texture analysis is a complex task itself and many works have been reported, mainly in the aspect of texture-based image segmentation [8]. However, the considered case is not identical to the segmentation task. Here, the aim is to define a suitable domain for color propagation, and we have found textural features very helpful for this purpose. Among the existing colorization methods, textural features have been already exploited for color transfer methods [5, 9]. However, only simple texture descriptors are used there, which may be helpful in some cases, but does not guarantee the distinctiveness between the regions marked with different scribbles.

3.1 Discriminative Textural Features

The color propagation domain should induce low costs between pixels belonging to a single scribble. On the other hand, the cost should be high, when the path crosses a boundary between areas marked with different scribbles. It is therefore important to find such image properties that would be uniform within a single scribble and different between the scribbles. It is worth noting that the most relevant properties may be different from case to case. For some images the luminance itself may be

sufficiently distinctive (e.g. in case of cartoons), while for others the variation in the gradient intensity may be relevant. In the work reported here we select the distinctive properties independently for every scribbled image using linear discriminant analysis (LDA). The analysis is performed over a set of simple image features extracted from pixels which belong to the scribbles. In this way we obtain the color propagation domain which is dynamically conformed to every specific case.

3.1.1 Linear Discriminant Analysis

Linear discriminant analysis [10] is a supervised statistical feature extraction method frequently used in machine learning. It finds a subspace defined by the most discriminative directions within a given training set of M -dimensional vectors classified into K classes. The analysis is performed first by computing two covariance matrices: within-class scatter matrix S_W :

$$S_W = \sum_{i=1}^K \sum_{\mathbf{u}_k \in K_i} (\mathbf{u}_k - \boldsymbol{\mu}_i)(\mathbf{u}_k - \boldsymbol{\mu}_i)^T, \quad (4)$$

and between-class scatter matrix S_B :

$$S_B = \sum_{i=1}^K (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \quad (5)$$

where $\boldsymbol{\mu}$ is a mean vector of the training set and $\boldsymbol{\mu}_i$ is a mean vector of the i -th class (termed K_i). Subsequently, the matrix $S = S_W^{-1}S_B$ is subjected to the eigen decomposition $S = \Phi \Lambda \Phi^T$, where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_M)$ is the matrix with the ordered eigenvalues along the diagonal and $\Phi = [\mathbf{v}_1 | \dots | \mathbf{v}_M]$ is the matrix with the correspondingly ordered eigenvectors as columns. The eigenvectors form the orthogonal basis of the feature space. Originally, the feature space has M dimensions, but only those associated with the highest eigenvalues have strong discriminative power, while the remaining can be rejected. In this way the dimensionality is reduced from M to m , where $m < M$.

After having built the m -dimensional feature space the feature vectors are obtained by projecting the original vectors \mathbf{u} onto the feature space: $\mathbf{v} = \Phi^T \mathbf{u}$. The similarity between the feature vectors is computed based on their Euclidean distance in the feature space.

3.1.2 LDA for Texture Analysis

In order to determine the discriminative features, first we calculate *basic image features* from every pixel. They are composed of: a) luminance, b) gradient intensity, c) local binary pattern, d) mean value and e) standard deviation computed in many kernels of different size, f) the difference between maximum and minimum values in the kernels, and g) the pixel value in the median filtered image. The basic features (d) – (g) were obtained for 15 kernel sizes ranging from 3×3 to 31×31 . Hence, every pixel x is described by an M -dimensional basic feature vector \mathbf{u}_x ($M = 63$ in the

presented case). The feature vectors of the scribble pixels are subsequently subject to LDA. Every scribble forms a separate class, so the analysis determines the most discriminative features between the scribbles for a given image. The feature vectors (\mathbf{v}) obtained using LDA are further termed *discriminative textural features* (DTF). The distance between any two feature vectors \mathbf{v}_1 and \mathbf{v}_2 in DTF space is computed as:

$$d = \sum_{i=1}^m (v_{1i} - v_{2i})^2. \quad (6)$$

During our experiments we observed that for the majority of analyzed cases it is sufficient to reduce the dimensionality of DTF vectors to $m = 2$. Also, we limit the number of the input vectors in each class to 100 so as to reduce the LDA training time. If a scribble contains more pixels, 100 of them are randomly selected. We have not observed any noticeable difference in the outcome compared to using all the scribble pixels, while the training time is definitely shorter.

3.2 DTF-Based Color Propagation Domain

After training, a projection matrix Φ is obtained and every pixel in the image is projected onto m -dimensional DTF space. An example of a scribbled image and its projection onto three leading LDA components representing the most discriminative textural features is presented in Fig. 1. The eigenvalues associated with these components are also given in the figure. It may be observed that these projections differentiate well between the areas marked with the scribbles. Also, 10 highest eigenvalues obtained for these two images are plotted in the figure. The values on the vertical axis are given relatively to the highest eigenvalue.

Fig. 2 shows an image with two scribbles (a) marked over flower and leafy background. The luminance of these pixels scaled from 0 to 100 is presented in (b) on the horizontal axis. It may be noted that although the flower pixels (F) are generally brighter than the background pixels (B), the luminance alone is not a discriminative feature here. The scribble pixels projected onto 2D DTF subspace are shown in (c). Here, two classes are well separated.

For every scribble a mean DTF feature vector is obtained and its DTF-distance (6) to every pixel in the image is computed in DTF space. In this way a DTF-distance map \mathbf{d}_i is obtained for every i -th scribble, which serves as a domain for determining the propagation paths. The local cost ρ from pixel x to y equals the y pixel value in the DTF distance map ($\rho(x, y) = \mathbf{d}_i(y)$). Examples of DTF-distance maps are presented in Fig. 3. These are the distances obtained for a landscape image (a) from the scribbles marked over: (b) sky, (c) volcano, and (d) ground.

It is worth noting that potentially the distance maps could be used directly for chrominance blending without determining the propagation paths. In this case, to obtain an i -th weight for a pixel x , the distance in DTF space $\mathbf{d}_i(x)$ would be used instead of the total path cost $C_i(x)$ in (3). Such approach does not guarantee continuity of the regions and therefore has not been utilized in the investigated cases.

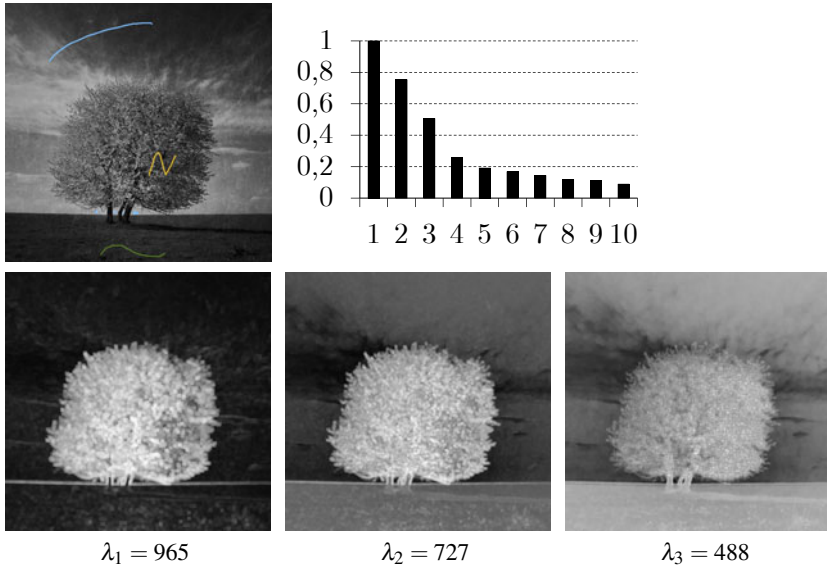


Fig. 1 Scribbled image and its projections onto three leading LDA components. A color version is available at http://sun.aei.polsl.pl/~mkawulok/cores2011/Figure_1.pdf.

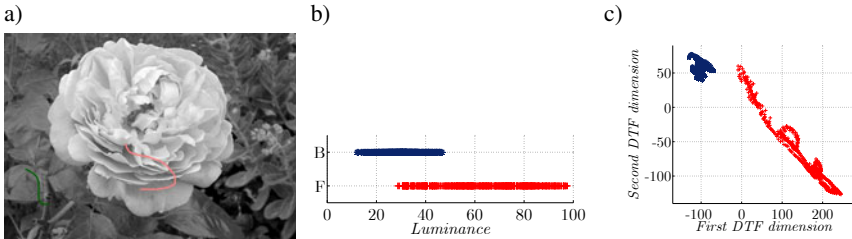


Fig. 2 Scribble pixels (a) projected onto luminance (b) and 2D LDA (c) subspaces. A color version is available at http://sun.aei.polsl.pl/~mkawulok/cores2011/Figure_2.pdf.

However, it may be useful for some other applications, e.g. for color transfer or video colorization.

The propagation paths are determined so as they follow the texture similar to that covered by the source scribble as long as possible. This is contrary to the conventional Yatziv's approach [2], with which the path is determined to minimize the gradient integrated in the propagation direction. An example of a difference between these two alternative approaches is given in Fig. 4. It shows the propagation paths leading from a scribble to a selected pixel obtained using two methods. The path determined using our method (b) does not leave the striped area, which allows to colorize the image correctly (c). The path obtained using a conventional method (d) shows that

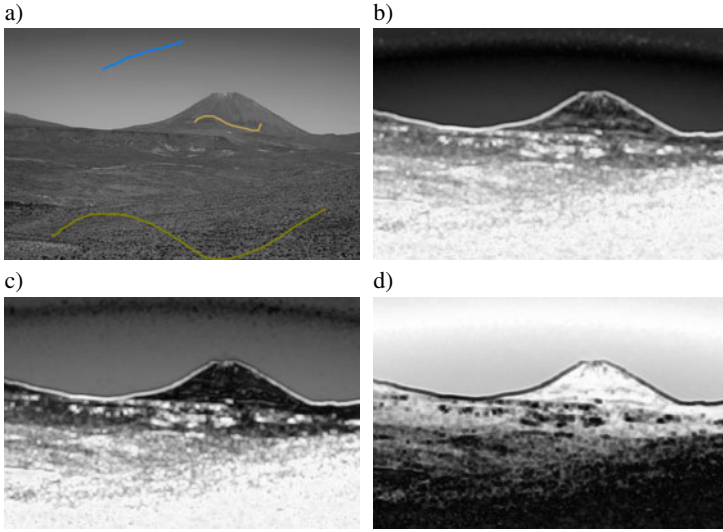


Fig. 3 Examples of DTF-distance maps (b-d) obtained for the scribbled image (a). A color version is available at http://sun.aei.polsl.pl/~mkawulok/cores2011/Figure_3.pdf.

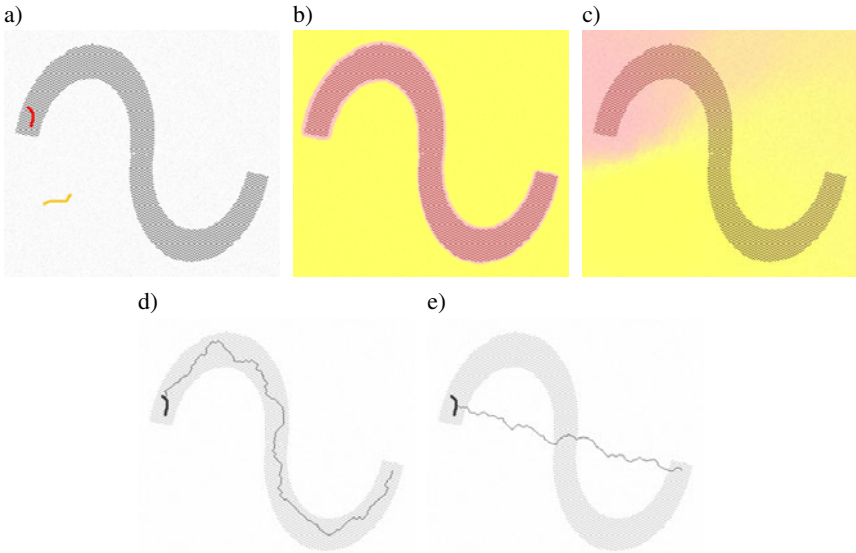


Fig. 4 Scribbled image (a), propagation paths and colorized image obtained using our (b,c) and Yatziv's approach (d,e). A color version is available at http://sun.aei.polsl.pl/~mkawulok/cores2011/Figure_4.pdf.

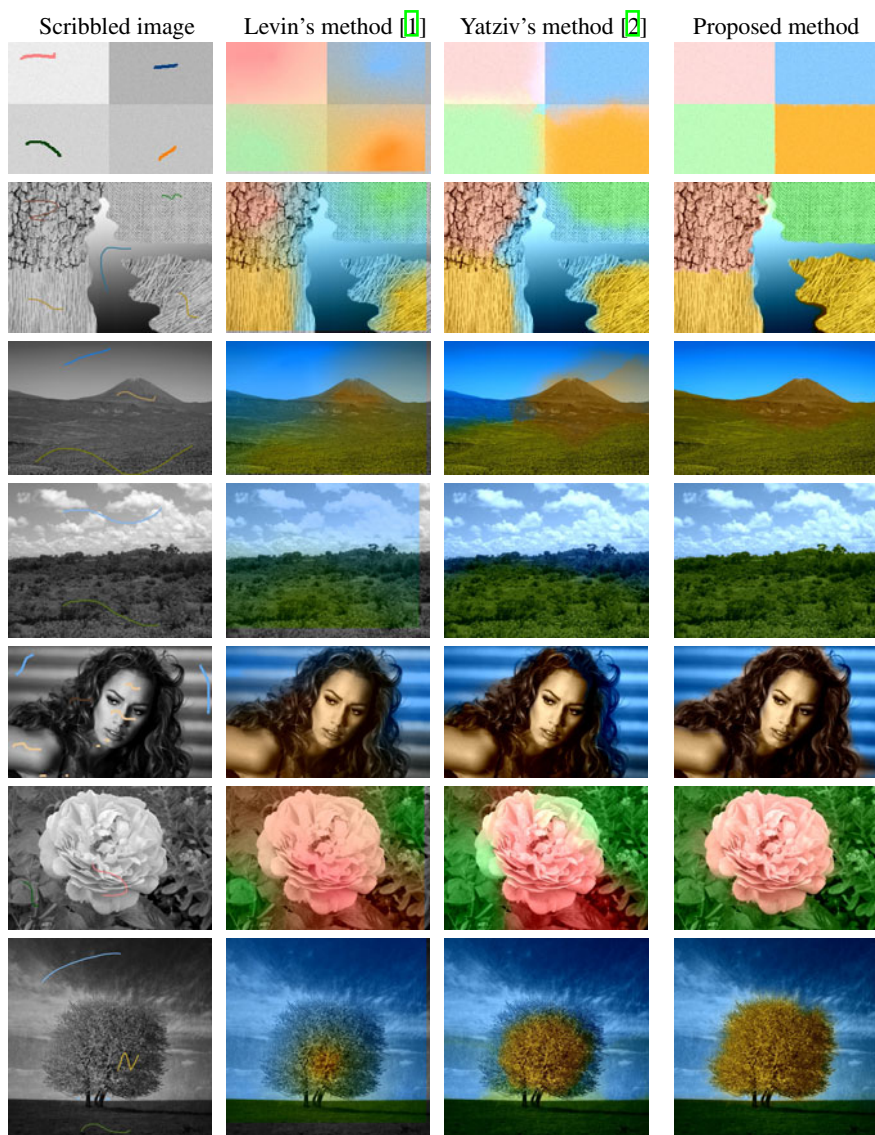


Fig. 5 Input image with scribbles and colorization results obtained with [1], [2] and our algorithm. A color version is available at http://sun.aei.polsl.pl/~mkawulok/cores2011/Figure_5.pdf.

the texture information is not taken into account during the propagation. This results in wrong colorization outcome (e).

In some cases the drawback of the presented method lies in the precision. At the boundaries of regions having different texture the pixels may be misclassified. Size

of such uncertain area depends on the kernel dimensions used for obtaining the basic textural features. This results in observing small halos at the region boundaries. We have approached this problem by determining a map of equal DTF distances. Then, in proximity of the equally distant pixels, the propagation domain is changed to the conventional one. This allows to remove the aforementioned effect. However, we are planning to develop a better solution during our future works.

4 Experimental Validation

We compared the proposed method with two well-established colorization techniques proposed by Levin [1] and Yatziv [2]. The first one is published in the form of MATLAB code and for the latter a Java applet is available to colorize a fixed set of images (for others we used our implementation). We evaluated the colorization on the basis of the visual result as this is a commonly adapted practice in image colorization.

Several examples of colorization result achieved using our method and alternative algorithms are given in Fig. 5. We validated our algorithm using three types of data: 1) artificial images to verify the theoretical assumptions (1st row), 2) semi-artificial images composed of various textures (2nd row), and 3) photographs to assess applicability of our approach. Our algorithm delivers the best visual effect in all of the presented cases, making it possible to colorize the textures with the largest precision. The halo effect mentioned in Section 3.2 can be observed around the flower (6th row) and the tree (bottom row). Also, small imprecisions can be noted for (1st, 2nd and 5th row), but they are definitely smaller than using the alternative methods. It is worth noting that the landscape photographs (3rd and 4th row) have been colorized perfectly using our method.

The time needed to complete the colorization process has not exceeded 20s for the tested cases. We have limited the number of pixels in every class to 100 and we use only two LDA dimensions. This assures that calculating the DTF space is not computationally expensive and takes a few percent of the time needed to colorize the image using conventional methods.

5 Conclusions and Future Work

In this paper we have presented a novel method for scribble-based image colorization, which uses the discriminant textural features domain for color propagation instead of the luminance channel. The experiments have shown that the proposed solution greatly improves the obtained results and decreases human assistance required in the colorization process.

Although the DTF domain works well for color propagation, its precision is limited at the region boundaries due to a large kernel size. Sometimes it results in unnatural halo effects which has been discussed in the paper and avoided in most cases using a simple technique. However, we are planning to develop a better solution to this problem during our future works. Moreover, we want to apply the proposed technique to color transfer and video colorization.

References

- [1] Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: SIGGRAPH, pp. 689–694 (2004)
- [2] Yatziv, L., Sapiro, G.: Fast image video colorization using chrominance blending. *IEEE Trans. on Image Proc.* 15, 1120–1129 (2006)
- [3] Lagodzinski, P., Smolka, B.: Digital image colorization based on probabilistic distance transform. In: Ruiz-Shulcloper, J., Kropatsch, W.G. (eds.) CIARP 2008. LNCS, vol. 5197, pp. 626–634. Springer, Heidelberg (2008)
- [4] Kim, T., Lee, K., Lee, S.: Edge-preserving colorization using data-driven random walks with restart. In: *IEEE ICIP*, pp. 1661–1664 (2009)
- [5] Welsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. *ACM Trans. Graph. (TOG)* 21, 277–280 (2002)
- [6] Ikonen, L., Toivanen, P.: Distance and nearest neighbor transforms on gray-level surfaces. *Pattern Rec. Lett.* 28, 604–612 (2007)
- [7] Heu, J., Hyun, D., Kim, C., Lee, S.: Image and video colorization based on prioritized source propagation. In: *IEEE ICIP*, pp. 465–468 (2009)
- [8] Zhang, J., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: a comprehensive study. *Int. J. of Computer Vision* 73, 213–238 (2007)
- [9] Lipowezky, U.: Grayscale aerial and space image colorization using texture classification. *Pattern Rec. Lett.* 27, 275–286 (2006)
- [10] Seber, G.: *Multivariate Observations*. Wiley, Chichester (1984)

Palette Generation in $L^*a^*b^*$ Colour Space Using ΔE

W.S. Mokrzycki and M. Tatol

Abstract. Existing colour spaces contain huge amount of colours describing their spaces. However, graphic devices are able to reconstruct only limited number of colours. This set of available reconstructed colours, called a colour palette, must be selected from a wide variety of possible colours in particular colour space. Device dependent colour spaces (for example RGB gamut) often uses many different palettes depend on specific needs. There are known many colour space quantization algorithms elaborated to that colour spaces, but we do not find palette for our needs. We try to find a palette for colour object identification tasks, so we consider $L^*a^*b^*$ colour space and ΔE difference.

Keywords: Perceptual colour spaces, colour difference measure, colour quantization, palette generation, $L^*a^*b^*$ colour space.

1 Introduction

1.1 Quantization Idea

Colour quantization can be viewed as a subset of the field of vector quantization. Stated simply vector quantization is the problem of selecting K vectors in some N -dimensional space to represent N vectors from that space where $K < N$ and the total error inferred by the quantization is minimal. Colour quantization is then vector quantization in a 3-Dimensional space (RGB, CIE, HLS etc.).

Colour image quantization process is often divide into four phases, [4].

- Sampling the original image for colour statistics.
- Choosing a colour map based on these statistics.
- Mapping the colours to their representation in the colour map.
- Quantizing and drawing the new image.

W.S. Mokrzycki · M. Tatol

Faculty of Mathematics and Informatics, University of Warmia and Mazury,
Żołnierska 47, Olsztyn, Poland

e-mail: [mokrzycki, mtatol}@matman.uwm.edu.pl](mailto:{mokrzycki, mtatol}@matman.uwm.edu.pl)

In general, algorithms for colour quantization can be divided into two categories: uniform and non-uniform quantization.

Uniform: the colour space is divided into equal size regions NR , $NR \leq K$ (see for example fig. [11](#)).

Non-Uniform: the colour space is divided depending on the distribution of colours in the image.

The remainder of this presentation will focus on 4 different algorithms: uniform, popularity, median cut and octree, [10](#).

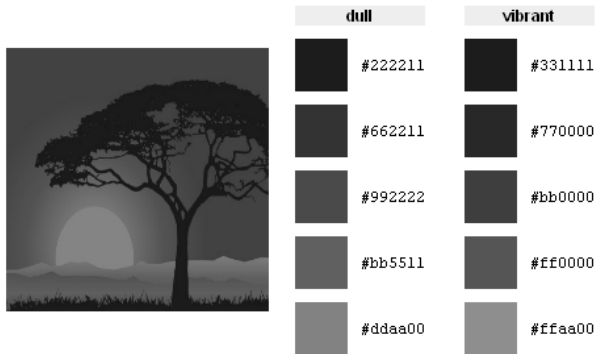


Fig. 1 Simple palette generator on DeGraeve webpage (RGB Hexadecimal), [12](#)

1.2 Uniform Quantization

In uniform quantization each axis of the colour space is considered independently and is divided into equal sized segments. The planes perpendicular to the axis passing through the division points define regions in the color space. The number of these regions is dependent on the scheme used for dividing the color space. One possible scheme is to divide the red and green axis into 8 segments each and the blue axis into 4 segments what results in 256 regions. Another possibility is dividing the red and blue into 6 and the green into 7 segments what results in 252 regions. Each one of these regions will produce a color for the color map, [5](#).

Once the color space has been divided, each of the original colors is then mapped to the suitable region. The representative colors of each region is the average of all the colors mapped to that region. Because each of the regions represents an entry in the color map, the same process for mapping the original colors to a region can be repeated. This algorithm is quick and easy to implement, but it does not yield very good results. Often region in the color space will not have any colors mapped to them, so color map entries are waste, [10](#).

1.3 Popularity Quantization Algorithm

Popularity algorithms are another form of uniform quantization. These algorithms divide the color space into much smaller so many more regions. One possible implementation is to divide the space into regions $4 \times 4 \times 4$ in size. The original colors are again mapped to the region they fall in. The representative color for each region is the average of the colours mapped to it. The color map is selected by taking the representative colors of the 256 most popular regions (the regions that had the most colours mapped to them).

These algorithms are also easy to implement and give better results than the uniform quantization algorithm. However, it takes slightly longer to execute and depending on the size of regions it can have a larger storage requirement. Depending on the image characteristics, it may not produce a good result. This can be said about all uniform sub-division schemes as the method for dividing the color space does utilize any information about the image.

1.4 Median Cut Algorithm

The premise behind median cut algorithms is to have every entry in the color map represent the same number of pixels in the original image. In contrast to uniform sub-division, these algorithms divide the color space based on the distribution of the original colors. The process is as follows, [2]:

1. Find the smallest box which contains all the colors in the image.
2. Sort the enclosed colors along the longest axis of the box.
3. Split the box into 2 regions at median of the sorted list.
4. Repeat the above process until the original color space has been divided into 256 regions.

The representative colors are found by averaging the colors in each box, and the appropriate color map index assigned to each color in that box.

Because these algorithms use image information in dividing the color space, this class of algorithms consistently produce good results while having memory and time complexity no worse than popularity algorithms.

1.5 Octree Algorithm

The principle of this algorithm is sequentially read of data. Every color is then stored in an octree of depth 8 (every leaf at depth 8 represents a distinct color). A limit of K (in this case $K = 256$) leaves is placed on the tree. Insertion of a color in the tree can result in two outcomes. If there are less than K leaves the color is filtered down the tree until either it reaches some leaf node that has an associated representative color or it reaches the leaf node representing its unique color.

If there are greater than K leaves in the tree some set of leaves in the tree must be merged (their representative colors averaged) together and a new representative

color stored in their parent. Authors [3] offer two possible criteria to be used in the selection of leaves to be merged. Reducible nodes that have the largest depth in the tree should be chosen first. They represent colors that lie closest together.

If there is more than one group of leaves at the maximum depth the algorithm could:

- Merge the leaves that represent the fewest number of pixels. This will help keep the error small.
- Reduce the leaves that represent the most pixels. In this case large areas will be uniformly filled in a slightly wrong color while maintaining detailed shadings.

Once the entire image has been processed in this manner the color map consists of the representative colors of the leaf nodes in the tree. The index of the color map is then stored at that leaf, and the process of quantizing the image is simply filtering each color down the tree until a leaf is hit.

As a limit is placed on the number of leaves in the tree this algorithm has a modest memory complexity, $O(K)$, compared to the median cut and popularity algorithms. The time complexity is more unclear. Authors [3] appreciate the search phase as being $O(N)$ where N is the number of pixels in the image. This is clearly best case behavior. The average case needs to address the complexity of the merging algorithm.

2 Description of $L^*a^*b^*$ Colour Space and ΔE 's

2.1 $L^*a^*b^*$ as *Perceptual Colour Space*

To create a formal description of human eye perception system, a colour spaces with linear eye perceptibility is needed. It means, that kind of colour space in which every dimension has the same linear modification. It is difficult, because of vision sense complexity and its abilities. The vision sense is dependent on environmental ambient, individual eye abilities or even age of person.

Perceptual colour spaces are designed to approximate human vision. They can be used to make accurate colour balance corrections. It is done by modifying A and B components, or to adjust the lightness intensity using the L. But perceptual colour spaces are larger than the gamut of computer displays, printers (using RGB and CMY colour spaces), or even human vision. A bitmap image represented as Lab requires more data per pixel to obtain the same precision as a RGB or CMYK bitmap, [7].

2.2 ΔE as *Difference Measure Factor in $L^*a^*b^*$ Colour Space*

Measuring colours is the first step in recognizing object property. These values, stored in the colour database, are being processed then. The most common operation is measure the perceptual difference. In perceptual uniform colour space it is

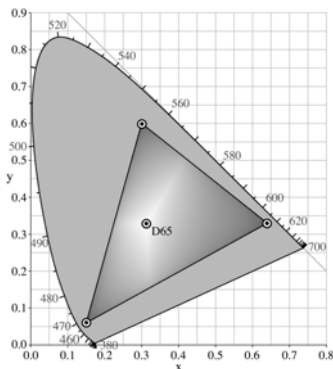


Fig. 2 RGB colour space

the distance vector called as ΔE . His value determines the difference between two measures. Depending on requirements type additional operations must be done.

Mathematical description of colour space presents colours as a group of numbers (see fig. 2). To find the distance between two random colours only an equation (on ΔE) is needed. If a colour space is linear, the measure reduces to a simple distance vector calculation. As well $L^*a^*b^*$ has closest human perception capabilities. During latest years there are few ΔE formulas were proposed, according to strict eye-like perception algorithm, [7].

3 Casting Range of $L^*a^*b^*$ Colour Space Quantization

Sometimes colour palette must be converted to destination colour space. Different colour spaces may have lower amount of colour or conversion process may fail - result value can be out of range of destination colour space. Also during preparations to begin a quantization process, some additional informations must be taken. If the quantization must take place over whole input colour space, in most colour space only extreme values are needed to set the borders. For example in RGB [9], tab. 1.

Table 1 Localization RGB space in $L^*a^*b^*$ space

Component	Value	R	G	B
L_{min}	0,00	0	0	0
L_{max}	99,98	1	1	1
A_{min}	-86,14	0	1	0
A_{max}	98,29	1	0	1
B_{min}	-107,84	0	0	1
B_{max}	94,50	1	1	0

If the range of input colour space is as well narrow down, it has to be taken into consideration, too. In non-uniform quantization the range of input space is marked on the $L^*a^*b^*$ colour space as a base to quantization process. Some of less popular colour (i.e. ProPhoto RGB developed by Kodak) spaces have values outside of natural $L^*a^*b^*$ colour space so it has to be kept in mind for further processing operations. Sample range of standardised RGB on $L^*a^*b^*$ gamut is shown on fig. 2.

4 Generation of Uniform and Non-uniform Palette

Uniform palette generation is broken into equal sized regions. This process of division is not dependent of any additional information. The important is the colour space range that will be taken into algorithm. This process is simple and give us fixed amount of colours, perceptually distant from each other. In this process ΔE is used:

$$\Delta E = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \quad (1)$$

This Delta E (1974) is the simplest method to measure the difference. There are more precise delta's (dE:CMC or dE2000) but they are also much more complex in calculations and results are almost equal in RGB range area of $L^*a^*b^*$ colour space [7]. If more precise calculations are needed, other delta's may be applied.

In non-uniform generation additional information must be taken from the consideration.

5 Conclusion

It was briefly described the problem of palette quantization in $L^*a^*b^*$ colour space. This method allow to generate more precise palette according to perceptual human eye properties. It can be very useful especially in that branch of industry, where colour sometimes must be compared by human (ie. varnishing). Usage of this colour space is strictly combined with ability of devices properties. In other way it can be convert to more popular or device-dependent colour space. Palette generation using $L^*a^*b^*$ is much more expensive in computations. Regardless this quantization is a single isolated operation. Generated palette can be then used multiple.

References

- [1] Bouman C.: Perceptually Uniform Colour Spaces. Digital Image Processing, June 26 (2007)
- [2] Furht, B. (ed.): Encyclopedia of Multimedia A-Z. Springer, Heidelberg (2008)
- [3] Gervautz, M., Purgathofer, W.: RISS - Ein Entwicklungssystem zur Generierung realistischer Bilder. Visualisierungstechniken und Algorithmen, pp. 61-79 (1988)

- [4] Heckbert, C.: Colour Image Quantization for Frame Buffer Display. *Computer Graphics* 16, 297–303 (1982)
- [5] Hill Jr., F.S.: *Computer Graphics*. Macmillan Publishing Co., Basingstoke (1990)
- [6] Hunter, S.R.: Photoelectric Colour-Difference Meter. *JOSA* 38(7), 661 (1948)
- [7] Mokrzycki, W.S., Tatol, M.: Perceptual difference in Lab Colour space as the base for object colour identification. In: *IPCC*, p. 382 (2009)
- [8] Orchards, M., Bouman, C.: Colour Quantization of Images. *IEEE Trans. on Sig. Proc.* 39(12), 2677–2690 (1991)
- [9] Henryk, P.: Selected issues of colour image processing. *ZN* 1742, 37 (2006)
- [10] Segenchuk, S.: An Overview of Colour Quantization Techniques. *CS* 563, May 5 (1997)
- [11] Bloomberg, D.S.: Color quantization using octrees,
<http://www.leptonica.com/papers/colorquant.pdf>
- [12] <http://www.degraeve.com/>
- [13] <http://www.brucelindbloom.com/>

Colour Image Quality Assessment Using the Combined Full-Reference Metric

Krzysztof Okarma

Abstract. In the paper the application of the combined image quality metric for the assessment of colour images is discussed. Proposed technique belongs to the group of full-reference objective methods, which require the exact knowledge of the reference image but ensure high universality and independence on the image contents. The combined metric discussed in the paper is based on three recently proposed approaches: Multi-Scale Structural Similarity, Visual Information Fidelity and RSVD metric utilising the Singular Value Decomposition. The verification of its linear correlation with subjective quality evaluations has been performed using two publicly available colour image databases: LIVE and TID2008.

Keywords: colour image quality assessment.

1 Introduction

Automatic image quality assessment is currently one of the most rapidly developing areas of image processing and analysis. Modern image quality metrics are also applied for the video files [25], usually using simple frame-by-frame approach. Starting from the Universal Image Quality Index [22] further extended into Structural Similarity [23], many approaches have been proposed. The main reason of such progress, apart from the growth of the possibilities of fast computing, e.g. using GPGPU programming techniques, is poor correlation of the classical Mean Squared Error and similar metrics [4, 5] with subjective image quality evaluation. Nevertheless, most of the metrics, even those recently proposed, do not use any colour information limiting the computations to the greyscale versions of the images. The

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

extension of such metrics into multichannel ones is not straightforward, similarly as e.g. vector median filtering, but the proper choice of the colour space may be very important as well as the choice of a specific quality metric. Considering the various properties of some recently proposed image quality assessment techniques, each of them has its specific properties and is more or less sensitive to specified types of distortions. An interesting idea for a metric well correlated with subjective scores, may be a nonlinear combination of some recent image quality assessment methods, proposed and verified for greyscale images in one of the earlier papers [11].

2 Recent Trends in Image Quality Assessment

The first step towards modern image quality assessment without any sophisticated model of the Human Visual System (HVS) has been the Universal Image Quality Index proposed in 2002 [22]. Despite its disadvantages, mainly related to its instability for dark or low variance regions of the image, it has become an impulse for many researchers, which has allowed the dynamic progress in this area of image analysis. Its further extension into the Structural Similarity (SSIM) is probably the most popular image quality assessment method, used also for the motion estimation [26] and modifications of lossy compression algorithms such as H.264/AVC [27].

The overall SSIM metric for the whole image can be computed as the mean of the local values calculated using the sliding window approach (11×11 pixels Gaussian window is the default window) according to the simplified formula [23]:

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

where C_1 and C_2 are small values preventing the possible division by zero. The default values suggested by the SSIM inventors are $C_1 = (0.01 \times 255)^2$ and $C_2 = (0.03 \times 255)^2$. The symbols \bar{x} and \bar{y} denote the mean values of the original and distorted image respectively, σ_x^2 , σ_y^2 and σ_{xy} stand for the respective variances and the covariance for the currently analysed fragments of both images. The default window can be changed, especially its size, especially depending on the resolution of images [10].

An interesting extension of this metric is the multi-scale version [24] using the dyadic pyramid. The luminance (l), contrast (c) and structure (s) factors are calculated for different scales and weighted using the specified exponents values. The final value of the MS-SSIM metric can be calculated as

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

where M stands for the highest scale obtained after $M - 1$ iterations of low-pass filtration and downsampling by a factor of 2.

Another metric proposed by the same group of researchers is the Visual Information Fidelity (VIF) metric [17, 18], which has two general versions. The basic

one utilises the wavelet decomposition and the simplified one operates in the pixel domain. Both of them are well correlated with subjective perception of some typical image distortions. The definition of this metric is

$$VIF = \frac{\sum_{j=0}^S \sum_{i=1}^{M_j} I(c_{i,j}; f_{i,j})}{\sum_{j=0}^S \sum_{i=1}^{M_j} I(c_{i,j}; e_{i,j})}, \quad (3)$$

where S stands for the number of sub-bands (or scales), M_j stands for the number of blocks at j -th sub-band (scale in the pixel domain) and $I(x; y)$ denotes the mutual information between x and y . The numerator and denominator are interpreted as the information that vision extracts from the distorted and reference images. It is assumed that c denotes a block vector at a specified location in the reference image, e is the perception of block c by a human viewer with additive noise n , and f is the perception of distorted block c [18].

Another promising idea for the image quality assessment is the application of the Singular Value Decomposition (SVD). The first attempt has been the metric proposed by Eskicioglu [19, 20] and in some of the recent publications the reflection factors [8] and R-SVD metric [9] have been discussed. Those two metrics are quite similar and better results can be achieved using the R-SVD metric, which can be computed as

$$\text{R-SVD} = \left(\frac{\sum_{i=1}^m (d_i - 1)^2}{\sum_{i=1}^m (d_i + 1)^2} \right)^{\frac{1}{2}}, \quad (4)$$

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

$$\hat{U}_i = \begin{cases} 0 & \text{if } \hat{s}_i = 0 \\ \hat{A} \cdot V_i / \hat{s}_i & \text{otherwise} \end{cases}, \quad (5)$$

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

3 Verification

Each newly proposed objective image quality assessment method requires some tests in order to determine its correlation with the way that humans perceive various distortions which may be present in the image. For this purpose an image database

containing the subjective quality scores is necessary. A typical approach is using the Mean Opinion Score (MOS) or Differential MOS values for the numerical expression of an average quality score for each image after the subjective assessment by numerous observers. The MOS values are obtained as the mean values of the scores derived by a number of observers which assess the quality of displayed images using the double stimulus approach. The most commonly used subjective quality scale is from 1 to 5 according to the recommendations of the ITU for such subjective tests.

Some of such databases contain a set of greyscale images, often with limited number of distortion types. A good example may be the Wireless Image Quality database (WIQ) consisting of 7 greyscale reference and 80 distorted images [2] containing only the distortions specific for wireless transmission. Another one is the Toyama database (known also as MICT) [13] containing 168 images obtained from 14 originals by JPEG and JPEG2000 compression. Another ones are: the IRC-CyN/IVC database [6] with 235 distorted images by 4 types of distortions, the LIVE database (779 distorted images with five types of distortions) [16] and Tampere Image Database (TID2008) containing 1700 images with 17 types of distortions [14]. Since two last databases are the most representative ones (many files and distortion types), they have been chosen for the experiments described in the paper.

The correlation of a given quality metric and subjective scores for a specified database can be expressed as Pearson's linear correlation coefficient, Spearman Rank Order Correlation Coefficient (SROCC), Kendall correlation, or the Root Mean Squared Error (RMSE) as well as using the outlier ratio. Since all those measures usually lead to the same conclusions, only the Pearson's correlation coefficients obtained for the proposed metric are presented in the paper.

Calculating the correlation coefficients all the DMOS/MOS scores for the reference (undistorted) images have been removed from the analysis, since they are predictable and would cause the unjustifiable increase of the correlation coefficient's values.

4 Combined Colour Image Quality Metric and Experimental Results

In order to obtain a metric which is highly correlated with subjective scores the nonlinear combination of three metrics has been proposed [11], which is expressed as:

$$\text{CQM} = (\text{MS-SSIM})^7 \cdot (\text{VIF})^{0.3} \cdot (\text{R-SVD})^{-0.15} . \quad (6)$$

The values of the exponents have been chosen as near optimal, leading to the linear Pearson's correlation coefficient equal to 0.86 for the whole TID2008 database and 0.7214 for the LIVE database. Nevertheless, those results have been obtained for the images converted to greyscale regardless of the presence of colour information. Such values are relatively high but in some publications one can find the correlations even over 0.9. Unfortunately such results cannot be compared directly, since many authors utilise the logistic function for nonlinear mapping suggested by the Video

Quality Experts Group (VQEG) [21]. The main drawback of such approach are the significant differences between the values of the coefficients of the logistic function obtained for each type of distortions present in a given database, as well as for various databases. In such case their proper choice requires additional optimization [15], seriously limiting the universality of such technique.

The same approach can also be used for the colour images separately for the RGB channels or using some other colour spaces. The experiments described in the paper have been conducted for the most popular colour spaces: RGB, YCbCr, HSV and CIE L*a*b* using three image quality metrics and their proposed combination. The results obtained for LIVE database (release 2 with realigned DMOS values) are presented in Table 1. Analysing the results a strong influence of the colour space and chosen channel on the results can be noticed. Proposed metric performs similarly as VIF for many channels. The most interesting results have been obtained for those two metrics using the luminance channel from CIE L*a*b* colour space. Such high linear correlation coefficients are much better than previously obtained (0.7214 for greyscale images as reported in [11]).

Table 2 demonstrates the results obtained for the TID2008 database. Presented results for the CIE L*a*b* luminance channel are slightly worse than for the greyscale images but illustrate the advantage of proposed combined index. Change of the exponents' values from [7 0.3 -0.15] to [5 0.5 -0.06] respectively increases slightly the correlation coefficient to 0.8487 for the L* channel. An interesting advantage of the combined quality metric for the RGB channels can be observed for the TID2008 database but for the images from the LIVE database it is not evident.

Highly linear relationship obtained for both databases using the L* channel is illustrated in Figs. 1 and 2 in comparison to Y luminance. The symbols used in the scatter plot for the LIVE database denote five distortion types.

Table 1 Pearson linear correlation coefficients obtained for the whole LIVE database

channel	MS-SSIM	VIF	RSVD	CQM
R	0.4891	0.7257	0.5812	0.7105
G	0.4711	0.7319	0.5871	0.7145
B	0.4752	0.6906	0.5755	0.6932
Y	0.4755	0.7311	0.5893	0.7282
Cb	0.4488	0.0482	0.3474	0.2482
Cr	0.4532	0.0095	0.3498	0.2405
H	0.5149	0.6511	0.2971	0.6989
S	0.4752	0.7532	0.5293	0.7781
V	0.4037	0.7544	0.5932	0.7509
L*	0.6671	0.9421	0.5726	0.9358
a*	0.5572	0.7207	0.4083	0.7643
b*	0.6138	0.7675	0.4549	0.7734

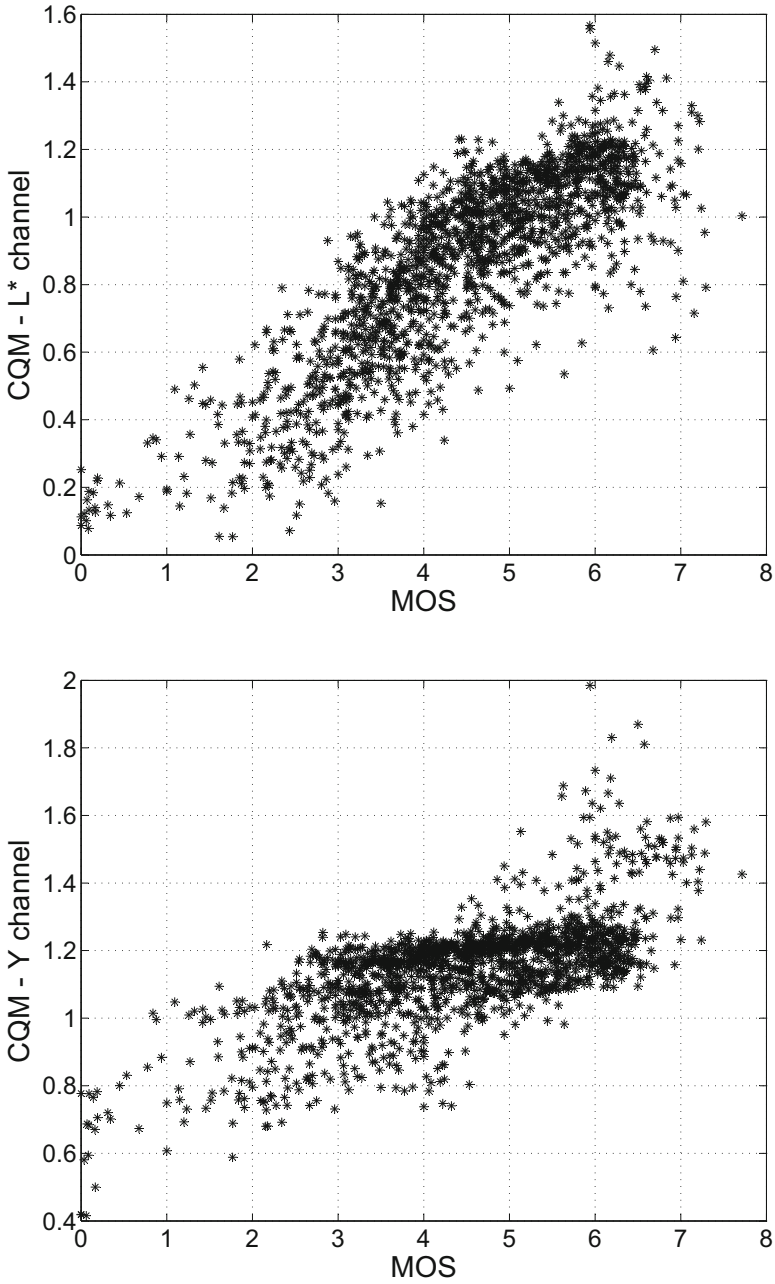


Fig. 1 Scatter plots of the CQM values calculated for L* and Y channels versus the MOS values for the 1700 images from the TID2008 database

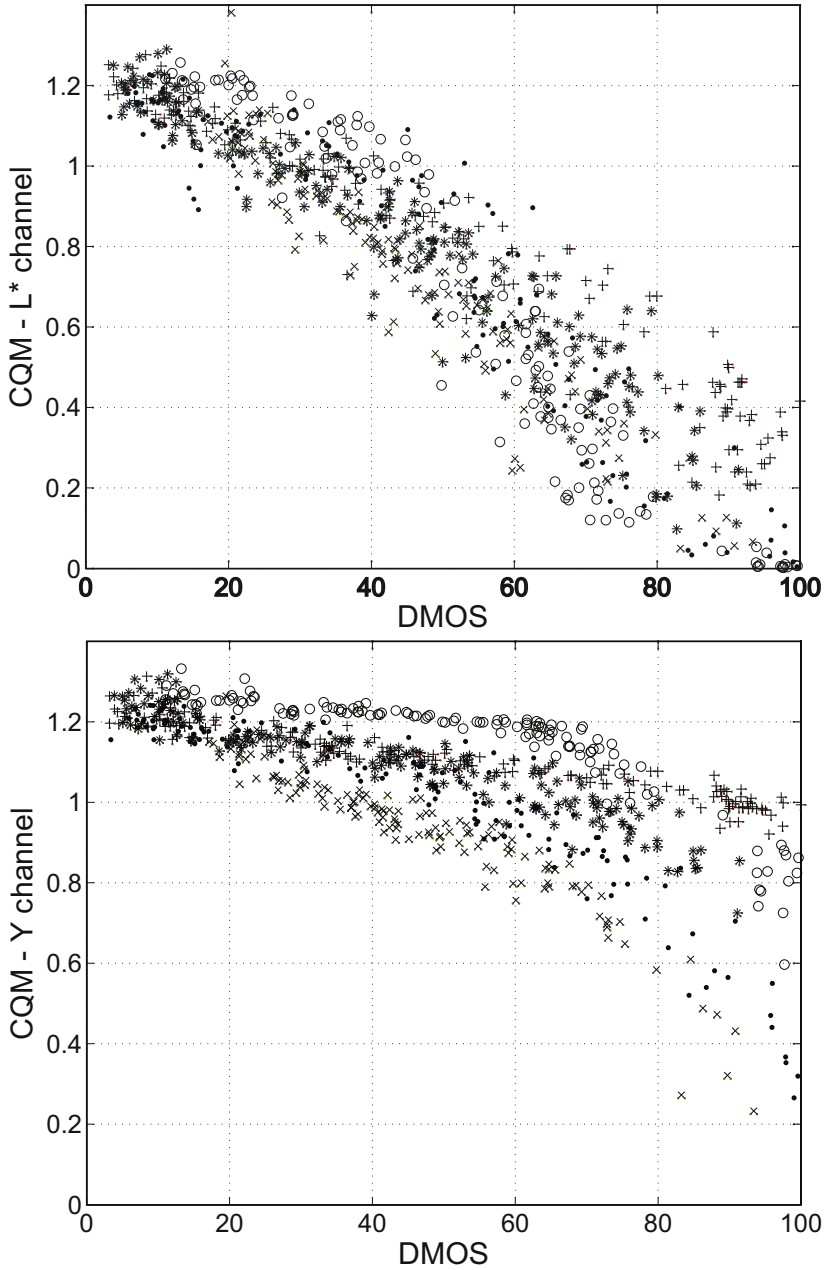


Fig. 2 Scatter plots of the CQM values calculated for L* and Y channels versus the DMOS values for the 779 images from the LIVE database

Table 2 Pearson linear correlation coefficients obtained for the whole TID2008 database

channel	MS-SSIM	VIF	RSVD	CQM
R	0.7335	0.5771	0.6654	0.7995
G	0.7391	0.5919	0.6685	0.8241
B	0.7033	0.5296	0.6638	0.7701
Y	0.0090	0.5975	0.6674	0.6811
Cb	0.3061	0.2764	0.5541	0.0752
Cr	0.2755	0.1005	0.5489	0.0671
H	0.1919	0.0197	0.4610	0.0493
S	0.1007	0.3102	0.5611	0.4502
V	0.0653	0.5949	0.6578	0.6700
L*	0.6824	0.8027	0.6604	0.8313
a*	0.3415	0.5768	0.5625	0.5100
b*	0.3601	0.5401	0.5635	0.4819

5 Conclusions and Future Work

The image quality assessment methods for colour images can be successfully applied for the luminance channels rather from the chrominance ones. The most promising results can be obtained using the CIE L*a*b* colour space together with combined quality metric. Nevertheless, some more complex tests, also using video quality assessment databases, should be performed in future work in order to examine the possibilities of using a combination of channels (e.g. L* but additionally supported by the chrominance) for the objective image and video quality assessment, as well as the application of recently proposed further modifications of the Structural Similarity, such as Gradient-based SSIM [1] and three-component SSIM [7].

Presented results, together with the analysis of the application of similar combined metric for the video quality assessment, presented recently [12], may be a starting point for further research towards a colour video quality metric even better correlated with Human Visual System. Possible directions for the extension of proposed approach can be the application of the Region Of Interest (ROI) approach [3] and saliency maps, which require the modelling of the most likely Region Of Interest based on the structural image analysis. For some types of images some existing solutions can be utilised e.g. face detection algorithms. Additionally, some temporal effects should be considered so the application of the motion estimation is necessary, similarly as in spatio-temporal filtering algorithms.

References

- [1] Chen, G.H., Yang, C.L., Xie, S.L.: Gradient-based structural similarity for image quality assessment In: Proc. Int. Conf. Image Processing, pp. 2929–2932 (2006)
- [2] Engelke, U., Kusuma, M., Zepernick, H.-J., Caldera, M.: Reduced-reference metric design for objective perceptual quality assessment in wireless imaging. Signal Processing: Image Communication 24(7), 525–547 (2009)

- [3] Engelke, U., Zepernick, H.-J.: Optimal Region-of-Interest based visual quality assessment. In: Proc. SPIE Human Vision and Electronic Imaging, vol. 7240 (2009)
- [4] Eskicioglu, A., Fisher, P., Chen, S.: Image quality measures and their performance. *IEEE Trans. Comm.* 43(12), 2959–2965 (1995)
- [5] Eskicioglu, A.: Quality measurement for monochrome compressed images in the past 25 years. In: Proc. Int. Conf. Acoust. Speech Signal Proc., pp. 1907–1910 (2000)
- [6] Le Callet, P., Autrusseau, F.: Subjective quality assessment IRCCyN/IVC database (2005), <http://www.irccyn.ec-nantes.fr/ivcdb/>
- [7] Li, C., Bovik, A.: Three-component weighted Structural Similarity index. In: Proc. SPIE Image Quality and System Performance, vol. 7242 (2009)
- [8] Mahmoudi-Aznavah, A., Mansouri, A., Torkamani-Azar, F., Eslami, M.: Image quality measurement besides distortion type classifying. *Optical Review* 16(1), 30–34 (2009)
- [9] Mansouri, A., Mahmoudi-Aznavah, A., Torkamani-Azar, F., Jahanshahi, J.A.: Image quality assessment using the Singular Value Decomposition theorem. *Optical Review* 16(2), 49–53 (2009)
- [10] Okarma, K.: Two-dimensional windowing in the structural similarity index for the colour image quality assessment. In: Jiang, X., Petkov, N. (eds.) CAIP 2009. LNCS, vol. 5702, pp. 501–508. Springer, Heidelberg (2009)
- [11] Okarma, K.: Combined full-reference image quality metric linearly correlated with subjective assessment. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010. LNCS (LNAI), vol. 6113, pp. 539–546. Springer, Heidelberg (2010)
- [12] Okarma, K.: Video quality assessment using the combined full-reference approach. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 2. AISC, vol. 84, pp. 51–58. Springer, Heidelberg (2010)
- [13] Parvez Sazzad, Z., Kawayoke, Y., Horita, Y.: Image quality evaluation database (2000), <http://mict.eng.u-toyama.ac.jp/mictdb.html>
- [14] Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., Battisti, F.: TID2008 - a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics* 10, 30–45 (2009)
- [15] Sendashonga, M., Labeau, F.: Low complexity image quality assessment using frequency domain transforms. In: Proc. IEEE Int. Conf. Image Proc., pp. 385–388 (2006)
- [16] Sheikh, H., Wang, Z., Cormack, L., Bovik, A.: LIVE image quality assessment database release 2 (2005), <http://live.ece.utexas.edu/research/quality>
- [17] Sheikh, H., Bovik, A., de Veciana, G.: An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Proc.* 14(12), 2117–2128 (2005)
- [18] Sheikh, H., Bovik, A.: Image information and visual quality. *IEEE Trans. Image Proc.* 15(2), 430–444 (2006)
- [19] Shnayderman, A., Gusev, A., Eskicioglu, A.: A multidimensional image quality measure using Singular Value Decomposition. In: Proc. SPIE Image Quality and Image Quality and System Performance, vol. 5294(1), pp. 82–92 (2003)
- [20] Shnayderman, A., Gusev, A., Eskicioglu, A.: An SVD-based gray-scale image quality measure for local and global assessment. *IEEE Trans. Image Proc.* 15(2), 422–429 (2006)
- [21] VQEG. Final report on the validation of objective models of video quality assessment (2003), <http://www.vqeg.org>
- [22] Wang, Z., Bovik, A.: A universal image quality index. *IEEE Signal Proc. Letters* 9(3), 81–84 (2002)

- [23] Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: From error measurement to Structural Similarity. *IEEE Trans. Image Proc.* 13(4), 600–612 (2004)
- [24] Wang, Z., Simoncelli, E., Bovik, A.: Multi-Scale Structural Similarity for image quality assessment. In: *Proc. 37th IEEE Asilomar Conf. on Signals, Systems and Computers* (2003)
- [25] Winkler, S.: *Digital video quality - vision models and metrics*. Wiley, Chichester (2005)
- [26] Yang, C.-L., Wang, H.-x., Po, L.-M.: A Novel Fast Motion Estimation Algorithm Based on SSIM for H.264 Video Coding. In: Ip, H.H.-S., Au, O.C., Leung, H., Sun, M.-T., Ma, W.-Y., Hu, S.-M. (eds.) *PCM 2007. LNCS*, vol. 4810, pp. 168–176. Springer, Heidelberg (2007)
- [27] Yang, C.-l., Leung, R.-K., Po, L.-M., Mai, Z.-Y.: An SSIM-optimal H.264/AVC inter frame encoder. In: *Proc. IEEE Int. Conf. Intel. Comp. and Intel. Syst.*, pp. 291–295 (2009)

Model-Based 3D Human Motion Capture Using Global-Local Particle Swarm Optimizations

Tomasz Krzeszowski, Bogdan Kwolek, and Konrad Wojciechowski

Abstract. We present an approach for tracking the articulated motion of humans using image sequences obtained from multiple calibrated cameras. A 3D human body model composed of eleven segments, which allows both rotation at joints and translation, is utilized to estimate the pose. We assume that the initial pose estimate is available. A modified swarm intelligence based searching scheme is utilized to perform motion tracking. At the beginning of each optimization cycle, we estimate the pose of the whole body and then we refine locally the limb poses using smaller number of particles. The results that were achieved in our experiments are compared with those produced by other state-of-the-art methods, with analyses carried out both through qualitative visual evaluations as well as quantitatively by the use of the motion capture data as ground truth. They indicate that our method outperforms the algorithm based on the ordinary particle swarm optimization.

1 Introduction

Vision-based tracking of human bodies is an important problem due to various potential applications like recognition and understanding human activities, user friendly interfaces, surveillance, clinical analysis and sport (biomechanics). The goal of body tracking is to estimate the joint angles of the human body at any time. This is one of the most challenging problems in computer vision and pattern recognition because of self-occlusions, high dimensional search space and high variability in human appearance. An articulated human body can be thought of as a kinematic chain with at least 11 body parts. This may involve around 26 parameters to describe the full body articulation. By building a mapping from configuration space to observation space, 3D model-based approaches rely on searching the pose space to find the body configuration that best-matches the current observations [1]. Matching

Tomasz Krzeszowski · Bogdan Kwolek · Konrad Wojciechowski
Polish-Japanese Institute of Information Technology,
Koszykowa 86, 02-008 Warsaw, Poland
e-mail: bytom@pjwstk.edu.pl

such complex and self-occluding model to human silhouette might be especially difficult in cluttered scenes. The major problems with 3D body tracking arise by reason of occlusions and depth ambiguities. Multiple cameras and simplified backgrounds are often employed to ameliorate some of such practical difficulties. However, the use of multiple cameras is connected with difficulties such as camera calibration and synchronization, as well as correspondence.

The particle filtering is widely used in human motion tracking [2] owing to ability of dealing with high-dimensional probability distributions. Given the number of parameters needed to describe an articulated model, the number of particles of that is required to adequately approximate the underlying probability distribution in the pose space might be huge. Recent work demonstrates that particle swarm optimization (PSO) can produce similar or even superior results over particle filtering due to capability exploration of the high dimensional search space [3].

Marker-less human motion capture has been studied since more than twenty years and is still a very active research area in computer vision and recognition. The advantage of marker-less technique is that it eliminates the need for the specialized equipment as well as time needed to attach the markers. Complete survey of marker-less human motion capture can be found in [4]. Despite huge research effort in this area, there have only been a few successful attempts to simultaneously capture video and 3D motion data serving as ground truth for marker-less motion tracking [2].

In this paper we present motion tracing results, which were obtained by a modified particle swarm optimization algorithm, together with analyses carried out both through qualitative visual evaluations as well as quantitatively by the use of the motion capture data as ground truth. The human body motion is modeled by a kinematic chain describing the movement of the torso/head, and both the arms as well as legs/feet. The tracking is done by particle swarm optimization with an objective function, which is built not only on cues like shape and edges but also on the segmented body parts. A global-local particle swarm optimization algorithm permits improved exploration of the search space as well as leads to better tracking, particularly regarding undesirable inconsistency in tracking legs and arms, resulting in swaps of such body parts, for instance, matching the right leg of the model to the left person's leg.

2 The Algorithm

2.1 Tracking Framework

The articulated model of the human body is built on kinematic chain with 11 segments. Such a 3D model consists of cuboids that model the pelvis, torso/head, upper and lower arm and legs. Its configuration is defined by 26 DOF and it is determined by position and orientation of the pelvis in the global coordinate system and the relative angles between the connected limbs. Each cuboid can be projected into 2D image plane via perspective projection. In this way we attain the image of the 3D model in a given configuration, which can then be matched to the person extracted through image analysis.

2.2 Person Segmentation

In most of the approaches to articulated object tracking, background subtraction algorithms are employed to extract a person undergoing tracking [5]. Additionally, image cues such as edges, ridges, color are often employed to improve the extraction of the person [6]. In this work we additionally perform the segmentation of the person's shape into individual body parts. In our approach, the background subtraction algorithm [7] is used to extract the person and to store it in a binary image. We model the skin color using 16×16 histogram in rg color space. The patches of skin color are determined through histogram back-projection. The skin areas are then refined using the binary image as mask. Given the height of the extracted person we perform a rough segmentation of the legs and feet.

2.3 PSO-Based Motion Tracking

Particle swarm optimization [8] is a global optimization, population-based evolutionary algorithm for dealing with problems in which a best solution can be represented as a point in n -dimensional space. The PSO is initialized with a group of random particles (hypothetical solutions) and then it searches hyperspace (i.e. R^n) of a problem for optima. Particles move through the solution space, and undergo evaluation according to some fitness function. Much of the success of PSO algorithms comes from the fact that individual particles have tendency to diverge from the best known position in any given iteration, enabling them to ignore local optima while the swarm as a whole gravitates towards the global extremum. If the optimization problem is dynamic, the aim is no more to seek the extrema, but to follow their progression through the space as closely as possible. Since the object tracking process is a dynamic optimization problem, the tracking can be achieved through incorporating the temporal continuity information into the traditional PSO algorithm. This means, that the tracking can be accomplished by a sequence of static PSO-based optimizations to calculate the best object location, followed by re-diversification of the particles to cover the possible object state in the next time step. In the simplest case, the re-diversification of the particle i can be realized as follows:

$$x_t^{(i)} \leftarrow \mathcal{N}(\hat{x}_{t-1}, \Sigma) \quad (1)$$

In the algorithm that we call global-local particle swarm optimization (GLPSO), at the beginning of each frame the estimation of the whole body pose takes place, see stage b) in Fig. 1. In the mentioned figure, the rectangular blocks represent state vectors with distinguished state variables of torso/head, arms, and legs/feet. To calculate such an estimate we initialize the particles using the estimated state \hat{x}_{t-1} in time $t - 1$, see stage a) in Fig. 1. In this stage the half of the particles is perturbed according to (1). In the remaining part of the swarm, the state variables describing the location of the pelvis are initialized using the linear motion, perturbed by normally distributed random motion with zero mean, whereas the remaining state variables are initialized using only normally distributed random motion with the

mean equal to the estimated state. Given the pose of the whole body, we construct state vectors consisting of the estimated state variables for torso/head and arms and state variables of the legs, which are constructed on the basis of the estimates of the corresponding state variables in time $t - 1$, see Fig. 1 and the arrow connecting the state variables in time $t - 1$ and time t . At this stage the discussed state variables are perturbed by normally distributed motion. Afterwards we execute particle swarm optimization in order to calculate the refined legs estimate, see stage d) as well as the right image in Fig. 1. Such refined state variables are then placed in the state vector, see stage e) in Fig. 1. The state variables describing the hands are refined analogously.

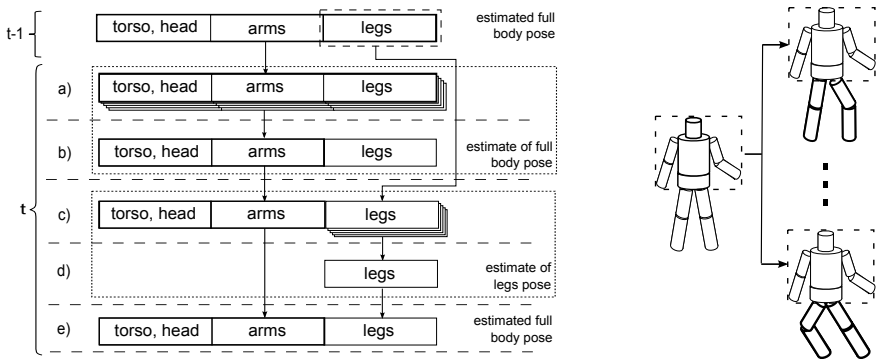


Fig. 1 Global-local particle swarm optimization for motion tracking (left), refinement of the legs configuration (right)

The fitness function of the PSO is determined on the basis of the following expression: $f(x) = w_1 f_1(x) + w_2 f_2(x) + w_3 f_3(x)$, where w_i stands for weighting coefficients that were determined experimentally. The function $f_1(x)$ reflects the degree of overlap between the segmented body parts and the projected segments of the model into 2D image. It is expressed as the sum of two components. The first component is the overlap between the binary image with distinguished body parts and the considered rasterized image of the model. The second component is the overlap between the rasterized image and the binary one. The larger the degree of overlap is, the larger is the fitness value. The function $f_2(x)$ reflects the degree of overlap of the edges. The last term, which reflects the overlap between arms and skin areas is not taken into account in the refinement of the pose of the legs.

3 Experiments

The algorithm has been tested in a multi-camera system consisting of four synchronized and calibrated cameras. The placement of the video cameras in our laboratory can be seen in Fig. 2. The cameras acquire images of size 1920×1080 with

rate 24 fps. Ground truth motion of the body is provided by a commercial motion capture (MoCap) system from Vicon Nexus at rate 100 Hz. The system uses reflective markers and sixteen cameras to recover the 3D position of such markers. The synchronization between the MoCap and multi-camera system is done through hardware from Vicon Giganet Lab.

The tracking performance of our motion tracking algorithm was evaluated experimentally in a scenario with a walking person, see Fig. 2. Although we focused on tracking of torso and legs, we also estimated the pose of both arms as well as of the head. The body pose is described by position and orientation of the pelvis in the global coordinate system as well as relative angles between the connected limbs. Figure 2 depicts the projected model and overlaid on the input images from four cameras. It is worth noting that the analysis of the human way of walking (gait) can be employed in various applications ranging from medical applications to surveillance. Gait analysis is currently an active research topic.

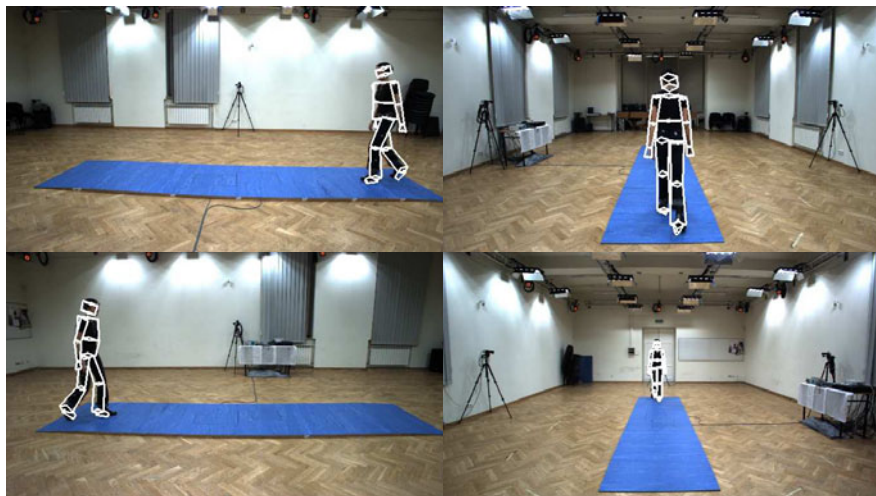


Fig. 2 Human motion tracking. Frame #20 in view 1 and 2 (upper row), in view 3 and 4 (bottom row)

Figure 3 illustrates some tracking results, which were obtained in one of the experiments. They were obtained using the global-local particle swarm optimization. The tracking of the whole body was done using 300 particles. In the pose refinement stage, the whole body is tracked using 200 particles, whereas 2×50 particles were used for tracking the legs as well as the arms. The tracking was achieved using 20 iterations in each stage mentioned above. The number of frames in the discussed sequence is equal to 180.

Figure 4 depicts some experimental results, which were achieved in another test sequence. The results were obtained by global-local PSO algorithm using the same number of the particles and iterations. The tracking was done on 150 images.

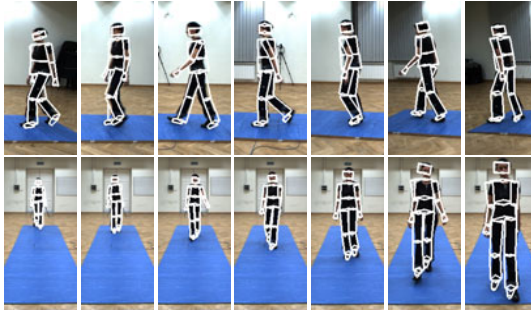


Fig. 3 Tracking results in a sequence 1. Frames #20, 40, 60, 80, 100, 120, 140 in view 1 (1st row) and in view 4 (2nd row)

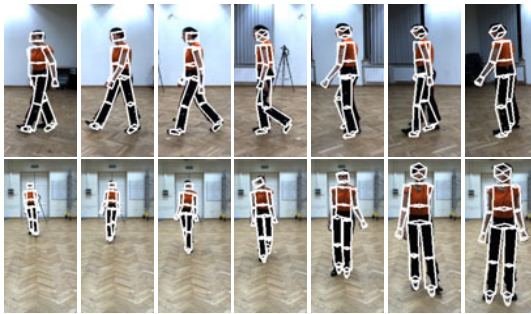


Fig. 4 Tracking results in a sequence 2. Frames #20, 40, 60, 80, 100, 120, 140 in view 1 (1st row) and in view 4 (2nd row)

The average Euclidean distance \bar{d}_i for each marker i was calculated using real world locations $m_i \in R^3$. It was calculated on the basis of the following expression:

$$\bar{d}_i = \frac{1}{T} \sum_{t=1}^T \|m_i(\hat{x}_t) - m_i(x_t)\| \quad (2)$$

where $m_i(\hat{x})$ stands for marker's position that was calculated using the estimated pose, $m_i(x)$ denotes the position, which has been determined using data from our motion capture system, whereas T stands for the number of frames. In Tab. [1](#) are shown the average distance errors \bar{d} for $M = 39$ markers. For each marker i the standard deviation σ_i was calculated on the basis of the following equation:

$$\sigma_i = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (\|m_i(\hat{x}_t) - m_i(x_t)\| - \bar{d}_i)^2} \quad (3)$$

The standard deviation $\bar{\sigma}$ shown in Tab. [1](#) is the average over all markers. From the above set of markers, four markers were placed on the head, seven markers on each arm, 6 on the legs, 5 on the torso and 4 markers were attached to the pelvis. Given

the estimated human pose and such a placement of the markers on the human body, the corresponding positions of virtual markers were calculated and then utilized in calculating the average Euclidean distance given by (2). The errors that are shown in Tab. 1 were obtained using frame sequences, which are depicted in Fig. 3 and Fig. 4. As we can observe, our GLPSO algorithm outperforms significantly the tracker that is based on the ordinary PSO.

Table 1 Average errors and standard deviations of the whole body tracking

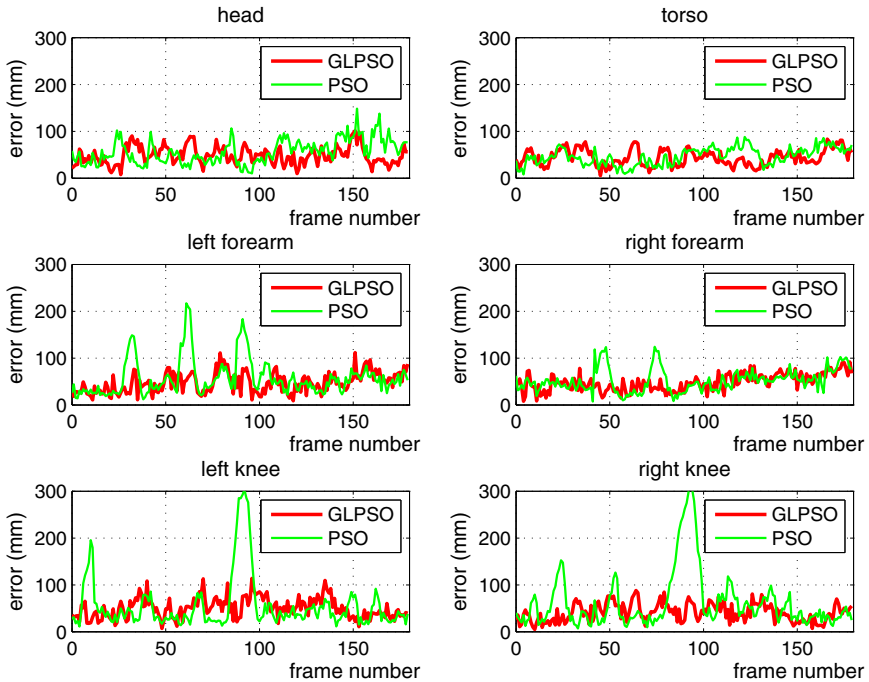
				Seq. 1		Seq. 2	
		#particles	#it.	error \bar{d} [mm]	$\bar{\sigma}$ [mm]	error \bar{d} [mm]	$\bar{\sigma}$ [mm]
PSO	100	10		79.41	50.19	82.25	53.48
	200	10		78.97	50.20	77.97	48.82
	300	10		74.54	46.24	77.84	50.31
	100	20		78.46	50.10	80.33	47.99
	200	20		72.32	44.28	77.29	49.17
	300	20		70.28	41.69	73.86	45.58
GLPSO	100	10		71.19	33.88	73.88	35.58
	200	10		67.17	27.83	66.36	25.58
	300	10		64.14	25.74	64.52	23.86
	100	20		68.03	28.31	67.01	27.72
	200	20		66.45	28.35	64.02	23.19
	300	20		62.46	23.77	62.59	22.45

For fairness, in all experiments we use the identical particle configurations. For the global-local PSO the sum of particles responsible for tracking the whole body, arms and legs corresponds to the number of the particles in the PSO. For instance, the use of 300 particles in PSO is equivalent to the use of 200 particles for tracking the full body, 50 particles for tracking the arms and 50 particles for tracking both legs in GLPSO. The use of 200 particles in the PSO corresponds to the exploitation of 100, 50 and 50 particles, respectively, whereas the use of 100 particles equals to utilization 60 particles for tracking the global configuration of the body, along with 20 and 20 particles for tracking hands and legs, respectively.

Table 2 shows the average errors and standard deviations of some markers, which are located on lower body limbs, i.e. the right knee and the right tibia. The discussed results were obtained on images from the sequence 1. As we can see, the difference between errors obtained by GLPSO and PSO algorithms is more significant in comparison to the results shown in Tab. 1. The difference is larger because in experiments, whose results are illustrated in Tab. 1 we considered several markers on torso as well as pelvis, which typically produce small errors, see also Fig. 5 and in consequence contribute towards more smoothed results.

Table 2 Average errors and standard deviations of lower body tracking

	#particles	#it.	PSO		GLPSO	
			error \bar{d} [mm]	$\bar{\sigma}$ [mm]	error \bar{d} [mm]	$\bar{\sigma}$ [mm]
right knee	100	20	68.08	59.88	50.13	33.18
	300	20	63.01	56.87	44.84	27.17
right tibia	100	20	95.57	84.72	70.40	44.12
	300	20	93.25	77.91	72.39	36.44

**Fig. 5** The errors [mm] versus frame number

In Fig. 5 we can observe a plot of the distance error over time for some body limbs. As we can observe, in some frames the PSO based algorithm produces considerably larger errors. The algorithm based on global-local PSO allows us to achieve superior results thanks to the decomposition of the search space.

In the last few years, there have only been a few successful attempts to simultaneously capture video and 3D motion data from the MoCap [2]. Hence, there are only few papers, which show the motion tracking accuracy by both qualitative visual analyses as well as quantitatively by the use of the motion capture data as ground truth. To our knowledge, the tracking accuracy presented above is as good as the

accuracy presented in [2] and in other state-of-the-art systems. However, instead of the annealed particle filter we utilize global-local particle swarm optimization. Moreover, our algorithm performs the segmentation of the person's shape into individual body parts. The global-local particle swarm optimization algorithm allows better exploration of the search space as well as allows us obtaining better tracking, particularly regarding undesirable swaps of legs and arms. The segmentation of the body into individual body parts contributes towards improved tracking, mainly due to better matching the individual parts of the model to corresponding body parts, extracted on the image.

The complete human motion capture system was written in C/C++. Currently, it requires an initial estimate of the body pose. The system operates on color images with spatial resolution of 960×540 pixels. The entire tracking process can be realized in approximately 1.1 sec. on a Intel Core i5 2.8 GHz using Open Multi-Processing (OpenMP), see Tab. 3. The image processing and analysis takes about 1 sec.

Table 3 Computation time on Intel Core i5 2.8 GHz

#particles	#it.	time [sec.]	
		1 core	4 cores
100	10	1.9	1.1
200	10	3.6	2.3
300	10	5.7	3.1
100	20	3.6	2.1
200	20	7.0	4.5
300	20	11.0	6.0

4 Conclusions

In this paper we presented an algorithm for tracking human pose using multiple calibrated cameras. The tracing is achieved by particle swarm optimization using both motion and shape cues. At the beginning the algorithm extracts person's silhouette and afterwards it segments the shape into individual body parts. A particle swarm optimization employs an objective function built on edges and such segmented body parts. We evaluated the algorithm through both qualitative visual analyses and quantitatively by the use of the motion capture data as ground truth. Experimental results show that algorithm achieves the tracking accuracy that is comparable to the accuracy produced by other state-of-art methods.

Acknowledgment. This work has been supported by the project "System with a library of modules for advanced analysis and an interactive synthesis of human motion" co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme - Priority Axis 1. Research and development of modern technologies, measure 1.3.1 Development projects.

References

- [1] Sidenbladh, H., Black, M., Fleet, D.: Stochastic tracking of 3D human figures using 2D image motion. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 702–718. Springer, Heidelberg (2000)
- [2] Sigal, L., Balan, A., Black, M.: HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *Int. Journal of Computer Vision* 87, 4–27 (2010)
- [3] Krzeszowski, T., Kwolek, B., Wojciechowski, K.: Articulated body motion tracking by combined particle swarm optimization and particle filtering. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010. LNCS, vol. 6374, pp. 147–154. Springer, Heidelberg (2010)
- [4] Poppe, R.: Vision-based human motion analysis: An overview. *Comp. Vision and Image Understanding* 108, 4–18 (2007)
- [5] Sminchisescu, C., Kanaujia, A., Li, Z., Metaxas, D.: Discriminative density propagation for 3D human motion estimation. In: *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. I, pp. 390–397 (2005)
- [6] Schmidt, J., Fritsch, J., Kwolek, B.: Kernel particle filter for real-time 3D body tracking in monocular color images. In: *IEEE Int. Conf. on Face and Gesture Rec.*, Southampton, UK, pp. 567–572. IEEE Computer Society Press, Los Alamitos (2006)
- [7] Arsic, D., Lyutskanov, A., Rigoll, G., Kwolek, B.: Multi camera person tracking applying a graph-cuts based foreground segmentation in a homography framework. In: *IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance*, pp. 30–37. IEEE Press, Piscataway (2009)
- [8] Kennedy, J., Eberhart, R.: Particle swarm optimization. In: *Proc. of IEEE Int. Conf. on Neural Networks*, pp. 1942–1948. IEEE Press, Piscataway (1995)

Gesture Data Modeling and Classification Based on Critical Points Approximation

Michał Cholewa and Przemysław Głomb

Abstract. Human-Computer Interaction (HCI) using natural gestures is one of the promising developments in User Interface technology. One of key issues in its design is reliable modeling and classification of gesture data. In this article, we present a method for abstraction of gesture movement information, by reducing it to a sequence of approximated critical points (locations and types). The sequence of such critical points has good feature extraction properties. We present the method, results of classification, and discussion of the properties of the representation based on example gesture dataset recorded with motion capture equipment.

1 Introduction

Gestures are a natural, easy to use and effective communication modality. They are now emerging as one of the most promising input data modalities for various types of user interfaces.

The key to designing a reliable gesture-based HCI is to develop effective and robust methods for classification of human motion data. This data is captured either with cameras—usually requiring further complex preprocessing to segment out acquisition effects (lighting, camera orientation, background, 2D to 3D conversion, etc)—or directly with motion capture equipment. Then, feature extraction and classification methods are needed to first to extract a representation maximizing inter-gesture while minimizing intra-gesture differences (i.e. the effect of making movement by different persons, speed) and second to make decision about recorded movement. Various approaches are used for this task, i.e. Vector Quantization and Hidden Markov Models [6]. For an in-depth review of methods, see [9].

We focus on designing a robust data representation and classification scheme for gesture classification of motion capture data. As related work shows polynomial

Michał Cholewa · Przemysław Głomb
Institute of Theoretical and Applied Informatics of PAS
Bałtycka 5, 44-100 Gliwice, Poland
e-mail: mcholewa@gmail.com, przemg@iitis.pl

representation is by far sufficient to be used as an approximation of motion readings. They are also quite easy to analyze as functions, with both derivatives and integrals fast and easy to calculate. We choose to represent the polynomials as a sequence of critical points (as opposed to coefficients), which are simpler to process and easier to work with (i.e. to notice errors of approximation).

In this article, we propose a method for representation and comparison of motion capture data. Section 2 presents the method, section 3 experimental results of classification with kNN on our gesture database; section 4 presents conclusions.

1.1 *Related Work*

Critical points are an established method of representing certain kinds of functions arising in data analysis. In [2], a problem of object shape description is analyzed; boundary line is decomposed to a sequence of arcs, which are segmented using ‘critical points’ (i.e. ‘corners’, that is points of discontinuity in the curvature, points of inflection, and curvature maxima). Various advantages of this approach are discussed. In [1] authors focus on extracting significant features from contours. Having assumed that the scale of a contour is unknown, the algorithm must focus on getting shape describing elements. With that in mind, authors present a method where critical points are used for curve shape description. Critical points are also an element of method of [10], who use them to describe noisy planar curves; a curve is segmented into convex and concave sections separated by inflection points.

There exists a number of methods of points approximation, as shown in review [7], which surveys methods of critical points extraction from digital data. Based on this analysis we have decided to develop a critical points extraction method based on polynomial approximation.

Polynomial approximation has been applied to processing of motion capture signals i.e. in [13] where it was used to approximate gait and was determined valuable, as small number of control points was needed to gain accuracy sufficient for medical sciences. Also, several attempts at recognition based on polynomial approximation have been proposed, i.e. [3] used dominant points of planar curves for approximating their shape, therefore achieving a method that can classify the curve regardless to its scale. Dominant points (points of intersection of shapes—a subset of critical points in polynomial case) were a valid way to store shape information.

2 Method

Our objective is the classification of a set of sensor readings (a recording of a person performing a gesture) using approximation of critical points of individual readings.

We have observed and empirically verified that each sensor readings can be reliably approximated by a polynomial. We represent the polynomial by a sequence of critical points (extrema, values at the end of the interval, and inflection points).

To compare two sensor readings we first select area of interest in both, focusing on the part that actually represents a gesture, and is not merely a result of a delay

or some slight hand movements that are hard to avoid when working with human subjects. The two areas of interest are compared using the weighted sum of relation between measure of area of interest, difference between polynomials on area of interest, difference between them beyond area of interest and value at the beginning and end of the gesture.

Knowing the difference between two sensor readings (as a real number) we can calculate distance between two gestures as a sum of differences of appropriate sensors. Having distance between two gestures we can apply a distance-based classifier (we use kNN) to assign gesture to a class.

2.1 Computing Critical Point Approximation for a Single Sensor

Each sensor data D is a set of two dimensional points $(t_i, y_i), i = 1, \dots, n$, with t_i time of registration, and y_i sensor reading value. The sensor data is subject to scaling and value normalization, performed as follows. Let

$$t_{min} = t_1 \quad t_{max} = t_n \quad \mu = \frac{\sum_{i=1}^n y_i}{n} \quad \sigma = \sqrt{\frac{\sum_{i=1}^n y_i - \mu}{n}}. \quad (1)$$

The preprocessing can be represented as the transformation $P : D \rightarrow D$

$$P((t_i, y_i)) = \left(\frac{t_i - t_{min}}{t_{max} - t_{min}}, \frac{y_i - \mu}{\sigma} \right). \quad (2)$$

In other words, the first coordinate is transformed into interval $\langle 0, 1 \rangle$, and the second is mean shifted and standard deviation normalized. Next, we calculate the critical points.

We consider a point to be a local maximum, if its y_i is highest in the interval $\langle t_i - \gamma, t_i + \gamma \rangle$. Value γ can be chosen depending on parameters of the sets. In our method, based on the analysis of the data, we used

$$\gamma = 3 \times \frac{\sum_{i=2}^n (t_i - t_{i-1})}{n - 1}. \quad (3)$$

Local minima are calculated in similar way.

Having a number of maxima and minima, we can use polynomial regression to approximate the dataset with a polynomial of given degree (determined by number of extrema) and calculate its inflexion points. We do this simply by using second derivative of polynomial being the effect of regression. Our final two points will be the values at the end of the interval.

After all calculations, our dataset D is converted into following data

1. Set V_D of points at the end of interval ($t_i = 0, t_i = 1$)
2. Set E_D of maxima and minima of the set (as above)
3. Set I_D of inflexion points of the polynomial approximating the dataset

Note that points in the set V_D are also local extrema. For future reference we can define D_{max} and D_{min} as maxima and minima in $V_D \cup E_D$.

2.2 Comparing Approximations

Selection of area of interest

During analysis of a dataset we must take into consideration that only a part of it represents the actual gesture. It is often the case that recording of movement starts too early or that there is additional reading before or after the gesture. Because of that, for dataset D we define ‘Area of interest’ (AOI) A_D as an interval within $\langle 0, 1 \rangle$ that contains data actually representing a gesture. For a purpose of the algorithm we define A_D as

$$A_D \in \left\{ \langle a, b \rangle : a, b \in V_D \cup E_D \right\}. \quad (4)$$

In other words A_D is an interval between two points that belong either to extrema set or endings.

When we compare datasets D_i and D_j we call two AOIs $A_{D_i} = \langle a_i, b_i \rangle$ and $A_{D_j} = \langle a_j, b_j \rangle$ relating if a_i and a_j are of the same type (maximum or minimum) and b_i and b_j are of the same type.

Transforming AOIs

Next, we scale AOI so they match - having two AOIs $\langle a, b \rangle, a = (t_1, y_1), b = (t_2, y_2)$ and $\langle c, d \rangle, c = (t_3, y_3), d = (t_4, y_4)$, we scale them so $t_1 = t_3$ and $t_2 = t_4$. Once we do that, we scale areas $\langle 0, t_1 \rangle, (t_1, t_2), (t_2, 1)$ and $\langle 0, t_3 \rangle, (t_3, t_4), (t_4, 1)$ accordingly. Having those set of transformed points we can approximate them with polynomials P_i and P_j .

Comparison

Once we have two datasets D_i and D_j with relating areas of interest A_{D_i} and A_{D_j} of sizes $|A_{D_i}| \leq 1$ and $|A_{D_j}| \leq 1$ and with polynomials approximating those datasets after transform (as in above paragraph) P_i and P_j we compute the difference between them as follows:

1. Scale $s_{ij} = \max \left\{ \frac{|A_{D_i}|}{|A_{D_j}|}, \frac{|A_{D_j}|}{|A_{D_i}|} \right\}$
2. Difference between points at the end of interval
 $p_{ij} = |y_1^{(i)} - y_1^{(j)}| + |y_n^{(i)} - y_n^{(j)}|$
3. Integral $q_{ij} = \int_0^1 |P_i(x) - P_j(x)| dx$
4. Final difference coefficient

$$d_{A_{D_i}, A_{D_j}}(D_i, D_j) = (s_{ij} - 1) + p_{ij} + 4q_{ij} \quad (5)$$

In other words we derive our difference from three values: scale between AOI in both datasets, difference in integral between transformed polynomials and difference between end points values. The coefficient before q_{ij} is selected considering the average value of s_{ij} , q_{ij} and p_{ij} , so they have the same weight towards final result. Right now those coefficients are calculated so that the end points difference of values and difference of integrals have on average the same magnitude, but it is a matter for further study. We use $s - 1$ as coefficient for scale to achieve difference between two identical gestures equal to zero.

Even though we calculate inflexion points they are at the moment not used as a part of classification algorithm, though they remain useful for eliminating redundant data from datasets.

Classification

Now we can compute the difference between datasets as

$$d(D_i, D_j) = \min_{AD_i, AD_j} \{d_{AD_i, AD_j}(D_i, D_j)\} \quad (6)$$

For classification, we use a weighted kNN classifier. By that we understand that the closest matching instance has a weight of k , next one $k - 1$ and so on, until finally $k - th$ has a weight of 1 and all further ones, 0. For each class we sum weight of all instances belonging to that class, and the gesture is classified to the one with the highest weight. If two classes have the same weight, we classify to the one with more representants with positive weight. When using weights, distance between closest neighbors and analyzed sample can be more important than actual number of closest neighbors.

3 Experiments

Data collection

For our experiment, we used base of 20 different type of gestures, each type represented by 15 instances - 3 people performing the gestures, each of them made the gesture 5 times (three with normal speed, then one fast following with one slow execution). The gestures are detailed in table 1. For discussion on gesture choice the reader is referred to [5]. The gestures were recorded with DG5VHand motion capture glove, containing 5 finger bend sensors (resistance type), and three-axis accelerometer producing three acceleration and two orientation readings. The time between each readings ($t_{i+1} - t_i$) was approximately 30ms.

Recognition experiments

Preliminary computations led us to conclusion that k for kNN classifier should be between 5 and 10. We decided to prepare a series of experiments with different k values, and weighted/unweighted classifier. Each gesture was compared to every

Table 1 The gesture list used in the experiments (note that it is a reduced version of the same set used in [12])

Name	Class ^a	Motion ^b	Comments
1 <i>A-OK</i>	symbolic	F	common ‘okay’ gesture
2 <i>Walking</i>	iconic	TF	fingers depict a walking person
3 <i>Cutting</i>	iconic	F	fingers portrait cutting a sheet of paper
4 <i>Showe away</i>	iconic	T	hand shoves away imaginary object
5 <i>Point at self</i>	deictic	RF	finger points at the user
6 <i>Thumbs up</i>	symbolic	RF	classic ‘thumbs up’ gesture
7 <i>Crazy</i>	symbolic	TRF	symbolizes ‘a crazy person’
8 <i>Knocking</i>	iconic	RF	finger in knocking motion
9 <i>Cutthroat</i>	symbolic	TR	common taunting gesture
10 <i>Money</i>	symbolic	F	popular ‘money’ sign
11 <i>Thumbs down</i>	symbolic	RF	classic ‘thumbs down’ gesture
12 <i>Doubting</i>	symbolic	F	popular Polish(?) flippant ‘I doubt’
13 <i>Continue</i>	iconic ^c	R	circular hand motion ‘continue’, ‘go on’
14 <i>Speaking</i>	iconic	F	hand portraits a speaking mouth
15 <i>Hello</i>	symbolic ^c	R	greeting gesture, waving hand motion
16 <i>Grasping</i>	manipulative	TF	grasping an object
17 <i>Scaling</i>	manipulative	F	finger movement depicts size change
18 <i>Rotating</i>	manipulative	R	hand rotation depicts object rotation
19 <i>Come here</i>	symbolic ^c	F	fingers waving; ‘come here’
20 <i>Telephone</i>	symbolic	TRF	popular Polish(?) ‘phone’ depiction

^a We use the terms ‘symbolic’, ‘deictic’, and ‘iconic’ based on McNeill & Levy [8] classification, supplemented with a category of ‘manipulative’ gestures (following [11]).

^b Significant motion components: T-hand translation, R-hand rotation, F-individual finger movement.

^c This gesture is usually accompanied with a specific object (deictic) reference.

other gesture in the set. We also wanted to check if transformation of accelerometers into velocity readings (numerical integration) has any effect on the results [4].

The results of the experiments are presented in table 2. We present the recognition ratio for each individual gesture - as it can be seen it can differ greatly between them which suggests the choice of gestures is important for the classifier. In first experiment, we assume k in kNN as 6 and we use weighted kNN. Second experiment, performed on the same set of data, incorporated transforming accelerometers to velocity readings. Apart from that we still used $k = 6$ and weighted kNN classifier.

¹ The finger bend and orientation are essentially position readings, while accelerometers can be viewed as second derivative of position. Also, integration lowers the number of extrema, thus speeding up algorithm’s execution.

Table 2 The results of gesture classification experiments: the success rate and errors (in parentheses – there were 15 input data points in all cases)

	Experiment 1 (wkNN, k=6)	Experiment 2 (wkNN, k=6, 'vel' ^a)	Experiment 3 (kNN, k=8, 'vel' ^a)	Experiment 4 (kNN, k=8)
1	93.3% (1)	100.0% (0)	100.0% (0)	93.3% (1)
2	100.0% (0)	100.0% (0)	100.0% (0)	100.0% (0)
3	100.0% (0)	93.3% (1)	86.7 % (2)	86.7 % (2)
4	60.0 % (6)	60.0 % (6)	60.0 % (6)	60.0 % (6)
5	93.3% (1)	100.0% (0)	100.0% (0)	100.0% (0)
6	93.3% (1)	93.3% (1)	93.3% (1)	93.3% (1)
7	86.7 % (2)	93.3% (1)	93.3% (1)	80.0 % (3)
8	93.3% (1)	53.3 % (7)	53.3 % (7)	86.7 % (2)
9	93.3% (1)	93.3% (1)	100.0% (0)	93.3% (1)
10	60.0 % (6)	40.0 % (9)	13.3 % (13)	26.6 % (11)
11	93.3% (1)	93.3% (1)	93.3% (1)	100.0% (0)
12	53.3 % (7)	46.7 % (8)	53.3 % (7)	53.3 % (7)
13	40.0 % (9)	33.3 % (10)	40.0 % (9)	46.7 % (8)
14	80.0 % (3)	66.7 % (5)	40.0 % (9)	60.0 % (6)
15	100.0% (0)	93.3% (1)	100.0% (0)	93.3% (1)
16	86.7 % (2)	86.7 % (2)	86.7 % (2)	86.7 % (2)
17	100.0% (0)	100.0% (0)	100.0% (0)	100.0% (0)
18	60.0 % (6)	60.0 % (6)	66.7 % (5)	60.0 % (6)
19	86.7 % (2)	80.0 % (3)	80.0 % (3)	93.3% (1)
20	80.0 % (3)	46.7 % (8)	60.0 % (6)	66.7 % (5)
total	82.3 % (53)	76.6 % (70)	76 % (72)	79 % (63)
total-2	86.2 % (37/270)	81.1 % (51/270)	81.5 % (50/270)	83.7 % (44/270)
total-4	89.6 % (25/240)	85.4 % (35/240)	85.8 % (34/240)	87.1 % (31/240)

^a Acceleration readings transformed to velocity readings.

In third experiment we used the same method of transforming data from accelerometers into velocity values, but this time we didn't use weighted kNN, but normal kNN classifier with $k = 8$. Finally we used unweighted kNN with $k = 8$ and no transformation.

3.1 Discussion

As there was no pre-selection of gestures we can safely assume that above results can roughly estimate basic efficiency of that algorithm. What must be noted,

however, is that this method can become even more accurate if we incorporate some criteria in selecting the gestures to match with - sufficiently 'far' by the means of our distance measure.

In case of our set of gestures, gestures 4, 10, 12 and 13 had worst results in all experiments, which suggests that they are candidates for elimination from the set. (For example gestures 'thumbs up' and 'thumbs down' are different in a small number of reading points, and 'cutting' gesture's finger movements are almost undetectable for the hardware used).

Another parameter that can influence the result is weight of three elements of distance measure (integral, scale and difference at the ends of interval). Manipulating those values could possibly result in improved efficiency.

We can also see that conversion from acceleration to velocity, even though seems irrelevant at first, produces noticeably worse results - this is the effect of having less extrema, so less choices of areas of interest.

4 Conclusion

This article presents a new approach on modeling and classification of motion capture gestures data based on critical points approximation. The sequences of estimated points are compared with a kNN classifier on a pre-prepared gesture database. The experimental results verify the reliability of the method.

This work has been partially supported by a Polish Ministry of Science and Higher Education project NN516405137 'User interface based on natural gestures for exploration of virtual 3D spaces'.

References

- [1] Bengtsson, A., Eklundh, J.-O.: IEEE Transactions on Pattern Analysis and Machine Intelligence 13(1), 85–93 (1991)
- [2] Freeman, H.: Pattern Recognition 10(3), 159–166 (1978)
- [3] Garcia, J.A., Fdez-Valdivia, J., Molina, R.: Signal Processing 43(1), 39–53 (1995)
- [4] Gavrikov, M.B., Misyurev, A.V., Pestryakova, N.V., Slavina, O.A.: Automation and Remote Control 67(2), 278–292 (2006)
- [5] GÅcomb, P., Romaszewski, M., Opozda, P., Sochan, A.: Choosing and modeling gesture database for natural user interface (2011) (submitted)
- [6] Lee, H.K., Kim, J.H.: IEEE Transactions on Pattern Analysis and Machine Intelligence 21(10), 961–973 (1999)
- [7] Li, Z.: The Cartographic Journal 32(2), 121–125 (1995)
- [8] McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. The University of Chicago Press, Chicago (1992)
- [9] Mitra, S., Acharya, T.: IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 37(3), 311–324 (2007)
- [10] Pikaz, A., Dinstein, I.: IEEE Transactions on Pattern Analysis and Machine Intelligence 16(8), 808–813 (2002)

- [11] Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., McCullough, K.E., Ansari, R.: ACM Transactions on Computer-Human Interaction 9, 171–193 (2002)
- [12] Romaszewski, M., GÅćomb, P.: The effect of multiple training sequences on HMM classification of motion capture gesture data accepted for 7th International Conference on Computer Recognition Systems CORES 2011 (2011)
- [13] Taylor, C.R., von Kinsky, B.R., Kirtley, C.: Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 5, pp. 2490–2493 (1998)

Evaluation of Illumination Compensation Approaches for ELGBPHS

Matthias Fischer, Marlies Rybnicek, and Christoph Fischer

Abstract. Various environmental conditions like pose variations, scale, noise and illumination changes cause matching problems for face recognition algorithms due to the fact that inappropriate data from images is extracted and consequently the recognition rate suffers. In the worst case, persons who should be accepted are rejected and vice versa. Enhanced Local Gabor Binary Patterns Histogram Sequence (ELGBPHS) is considered as an advanced and robust face recognition method. In this paper we evaluated if state-of-the-art illumination compensation approaches can further improve the performance of ELGBPHS. The paper outlines if it is worth to additionally implement preprocessing steps with the increasing complexity and cost. Therefore tests were performed to check if the recognition rate improves if applying preprocessing steps and adjusting essential parameters. Multi-Scale-Retinex, Histogram Equalization, 2D discrete Wavelet-Transformation and one approach combining Gamma Correction, Difference of Gaussian Filtering and Contrast Equalization (TT) were implemented and evaluated.

1 Introduction

In the practical field 2-dimensional face recognition systems are not considered to be working successfully until they achieve high recognition rates and keep the False Acceptance Rate (FAR) and False Rejection Rate (FRR) low. This is usually achieved by adding preprocessing stages for compensating several environmental conditions - like pose variations, scale, illumination changes and noise - before a face recognition algorithm is applied on an image. Those disturbing conditions can have substantial effects on the robustness of face detection and recognition systems [5]. In the worst case a proper person gets rejected by the system even though the

Matthias Fischer · Marlies Rybnicek · Christoph Fischer

St. Poelten University of Applied Sciences,

Institute of IT-Security Research, Matthias Corvinus-Strasse 15, 3100 St. Poelten, Austria

e-mail: [matthias.fischer,marlies.rybnicek}@fhstp.ac.at](mailto:{matthias.fischer,marlies.rybnicek}@fhstp.ac.at),

christoph.fischer@fhstp.ac.at

person is registered as an authorized person. In the proposed paper the focus lies on illumination changes in order to improve the recognition rate of Enhanced Local Gabor Binary Patterns Histogram Sequence (ELGBPHS) [20]. The main contribution of this paper is to answer the question whether preprocessing methods can further improve the recognition rate of ELGBPHS.

Face recognition algorithms can be roughly divided into holistic, feature-based and hybrid methods [21]. ELGBPHS [20] is a holistic approach which is based on Local Binary Patterns (LBP) [2]. In ELGBPHS a face image is convoluted with Gabor Wavelets. The resulting magnitude and phase maps are processed with LBP. Afterwards the maps are divided into sub-blocks, Histograms of the sub-blocks are built and by means of Histogram Intersection the similarity of faces is calculated. Increments of the Local Binary Patterns include Histograms of Gabor Phase Patterns (HGPP) [19] and Local Gabor XOR Patterns (LGXP) [18]. These three approaches are based on Gabor Wavelets which are convoluted with face images to get magnitude and phase maps of an image. The assignment of Gabor Wavelets for face recognition increases [12]. Not only holistic approaches exploit them. Popular local-based face recognition algorithms like the Elastic Bunch Graph Matching (EBGM) [3, 17] or Dynamic Link Architecture (DLA) [9] are based upon them too. Local-based methods use Gabor Wavelets to extract features like the eyes or the tip of the nose. In the training phase data around such characteristic points is extracted and used to create a so-called bunch graph which is a collection of all nodes and/or edges of the training images. By means of the bunch-graph a new graph can be found and placed over a new, unknown face. To sum it up, the result of the training phase is a so-called bunch graph which helps to find new image graphs in face images. The similarity of faces can be calculated by the comparison of the nodes and/or the edges of extracted face image graphs. Humans use both local and global features to recognize people [13]. Being aware of this fact hybrid approaches combine holistic and local-based methods to achieve good recognition rates [14].

Overall four preprocessing chains were implemented. All chains have two steps in common: Face Detection and Resize of an image.

First the Multi-Scale Retinex (MSR) Algorithm [7, 8, 11] is discussed. MSR splits the image into illumination and reflectance part. The aim is to adjust the illumination part that is mainly responsible for illumination changes. Further Colour Restoration is applied on the reflectance part and Gamma Correction on the illumination image. Because of the fact that the gamma value can be adapted within certain ranges tests were performed to determine a reasonable value for good illumination normalization.

In addition Histogram Equalization [1] as a well known efficient method is applied to enhance the image's contrast by spreading the grey values within a defined new range gaining new uniformly distributed grey values.

Moreover 2D discrete Wavelet-Transformation, an illumination approach in the wavelet domain, as proposed in [4] is described. To get low- and high-frequency coefficients the image is divided into four sub-bands using Daubechies-Wavelets. The coefficients are modified by means of different methods. The altered subimages are subsequently recombined to form the output image.

Least the approach proposed in [15] was implemented that combines Gamma Correction, Difference of Gaussian Filtering (DoG) and Contrast Equalization which counters the effects of illumination variations, local shadowing and highlights.

The remainder of this paper is organized as follows: Section 2 describes the face recognition algorithm ELGBPHS. Section 3 explains the applied preprocessing approaches for illumination compensation. Section 4 and 5 show experimental results and conclusion.

2 Face Recognition Algorithm

2.1 Enhanced Local Gabor Binary Patterns Histogram Sequence

In this paper preprocessing methods are evaluated with respect to the holistic face recognition algorithm Enhanced Local Gabor Binary Patterns Histogram Sequence (ELGBPHS) [20]. After convoluting face images with Gabor Wavelets and applying Local Binary Patterns (LBP) [2] on the resulting magnitude and phase maps, Histograms are calculated and by means of Histogram Intersection the similarity of face images is computed. For convolution 2dimensional Gabor Wavelets are used with the parameters orientation θ , frequency λ , phase φ , size σ and aspect ratio γ which define the form of the wavelet. The values for the parameters are taken from the master thesis of Bolme [3].

$$W(x, y, \Theta, \lambda, \varphi, \sigma, \gamma) = e^{-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}} \cos(2\pi \frac{x'}{\lambda} + \varphi) \quad (1)$$

$$x' = x \cos \Theta + y \sin \Theta$$

$$y' = -x \sin \Theta + y \cos \Theta$$

The result of the convolution are magnitude and phase maps that are processed with LBP and afterwards are divided into 10x10 (width x height) sub-blocks. We get 100 sub-blocks. The LBP operator processes all pixel values f_c by comparing it to its neighbourhood $f_p(0, 1, \dots, 7)$.

$$S(f_p - f_c) = \begin{cases} 1, & \text{if } f_p \geq f_c \\ 0, & \text{if } f_p < f_c \end{cases} \quad (2)$$

Afterwards each pixel value is calculated by using the factor 2^p as follows:

$$LBP = \sum_{p=0}^7 (S(f_p - f_c) 2^p) \quad (3)$$

From the resulting LBP images Histograms are built that can be compared by means of Histogram Intersection. For every face image 8000 Histograms (there are 40 magnitude and 40 phase maps for every sub-block) are built. By weighting the sub-blocks $W_{\theta, \lambda, r}^m$ one can emphasize important regions of a face like the eyes, the nose

or the mouth. Uninteresting parts of the face can be ignored by setting the weight to zero. So the biometric template of one face consists of histograms \mathfrak{R}^m and \mathfrak{R}^p of the magnitude and phase LBP images. To compare two faces one has to calculate the Histogram Intersection of the corresponding sub-blocks. This is summed up over all orientations, frequencies, regions and phases. Ψ denotes the intersection of two histograms and R the number of subregions of the image.

$$\mathfrak{R} = (\mathfrak{R}^m, \mathfrak{R}^p) \quad (4)$$

$$S(\mathfrak{R}_1, \mathfrak{R}_2) = S^m(\mathfrak{R}_1, \mathfrak{R}_2) + S^p(\mathfrak{R}_1, \mathfrak{R}_2) \quad (5)$$

$$S^m(\mathfrak{R}_1, \mathfrak{R}_2) = \sum_{\theta=0}^7 \sum_{\lambda=0}^4 \sum_{r=0}^{R-1} W_{\theta,\lambda,r}^m \Psi(H1_{\theta,\lambda,r}^m, H2_{\theta,\lambda,r}^m) \quad (6)$$

$$S^p(\mathfrak{R}_1, \mathfrak{R}_2) = \sum_{\theta=0}^7 \sum_{\lambda=0}^4 \sum_{r=0}^{R-1} W_{\theta,\lambda,r}^p \Psi(H1_{\theta,\lambda,r}^p, H2_{\theta,\lambda,r}^p) \quad (7)$$

The higher the resulting sum the similar the faces.

3 Preprocessing Approaches

Preprocessing steps can support the recognition process by compensating illumination changes or noise in images before a face recognition algorithm is applied on an input image. The goal is to remove disturbing data. In this paper different preprocessing approaches are evaluated with respect to the face recognition algorithm ELGBPHS [20].

3.1 Common Preprocessing Steps

All preprocessing chains have two steps in common which are performed at the beginning, namely Face Detection and Resize.

Face Detection: For Face Detection the HaarClassifier Cascade implemented in the OpenCV Library [6] is used. It is based on the Viola-Jones detector developed by Paul Viola and Michael Jones [16]. The detector is trained to find Haar-like features in images and is able to find faces. One major advantage is that the position of a person's eyes nearly matches perfectly. This is very useful for following steps. The result of Face Detection is a 200x200 (height x width) image of the face.

Resize: After Face Detection the image is resized to 100 x 100 (width x height) to further reduce the data amount and ensure the same input for all following preprocessing approaches.

3.2 Multi-scale Retinex

Multi-Scale Retinex (MSR) [7, 8, 11] is a method to compensate illumination changes in colour images. It is worth to mention that MSR is the only algorithm that works on colour images, all others process grey-scale images. Because the HaarCascade Classifier only works on grey-scale images, MSR is applied on the original images before Face Detection and Resize take place. Therefore the input image $S(x, y)$ is divided into illumination $L(x, y)$ and reflectance part $R(x, y)$.

$$S(x, y) = R(x, y) * L(x, y) \quad (8)$$

Colour Restoration and Image Stretching are performed additionally to further improve the resulting image. To get the illumination, the reflectance part is computed by the following formula:

$$R_i(x_1, x_2) = \sum_{k=1}^K W_k (\log I_i(x_1, x_2) - \log [F_k(x_1, x_2) * I_i(x_1, x_2)]) \quad (9)$$

$F_K(x, y)$ defines the neighbourhood function. The function is set manually, so a wide scope of filters can be used. For the evaluation a Gaussian filter was used. W_k represents the weight of the neighbourhood functions. As default value 1 is used and $k > 1$ is the number of neighbourhood functions. Five functions with different pixel numbers were used for MSR. Colour-Restoration is applied on the reflectance part to enhance the pixel values of the colour channels as given by the formula:

$$C_i(x, y) = \beta \log[\alpha I'_i(x, y)] = \beta \{ \log[\alpha I_i(x, y)] - \log[\sum_{i=1}^S (I_i(x, y))] \} \quad (10)$$

α is a gain constant and β controls the strength of nonlinearity. The first term represents the current colour channel and the second term the sum of the values over all colour channels at a given pixel position. To get the illumination part L , the input image is transformed to the logarithm domain and the reflectance part R is subtracted from the original image S . This step is necessary because the illumination part is essential for further illumination compensation. Gamma Correction is applied on the illumination part. Before applying Gamma Correction, L has to be transformed back to the spatial domain.

$$L' = W \left(\frac{L}{W} \right)^{\frac{1}{\gamma}} \quad (11)$$

W defines the highest value in the RGB space (255).

$$S' = L' * R = \left(\frac{L'}{L} \right) * S \quad (12)$$

Finally the new illumination image is multiplied with the colour-restored reflectance image generating the new output image named Bright Retinex Image. Gamma Correction is only accomplished on the original illumination image.

3.3 Histogram Equalization

Histogram Equalization (HE) [11] is a common and efficient method to increase the global contrast and to compensate illumination changes in an image. Therefore images are represented as histograms of their grey values.

3.4 Daubechies-Wavelet-Transformation

Daubechies-Wavelet-Transformation [4] is a method for normalizing an image in the wavelet domain with respect to illumination changes. Applying the 1-level Daubechies-Wavelet-Transformation on the whole image row-wise and column-wise leads to four low- and high-frequency sub-bands. The LL sub-band was processed by Histogram Equalization to enhance the contrast of the approximation coefficients, the other sub-bands were magnified by multiplying with the scalar 2 to enlarge the detailed coefficients. To reconstruct the image inverse Wavelet-Transform was used.

3.5 Gamma Correction, Difference of Gaussian Filtering and Contrast Equalization (TT)

In [15] an illumination normalization process is presented which counters the effect of illumination variations, local shadowing and highlights. It is based upon three stages, namely Gamma Correction, Difference of Gaussian (DoG) Filtering and Contrast Equalization. Gamma Correction enhances the dynamic range of dark or shadowed regions and concurrently compacts the range in bright regions. For our tests we used the recommended gamma value γ of 0.2 and for Difference of Gaussian Filtering the values $\sigma_0 = 1$ and $\sigma_1 = 2$. DoG is exploited because of its bandpass behavior. The last step namely Contrast Equalization rescales the image intensities to overcome extreme values produced by highlights and small dark regions. The pixel intensities are normalized in the following three stages where $\alpha = 0.1$ and $\tau = 10$:

$$I(x,y) \leftarrow \frac{I(x,y)}{(\text{mean}(|I(x',y')|^\alpha))^{\frac{1}{\alpha}}} \quad (13)$$

$$I(x,y) \leftarrow \frac{I(x,y)}{(\text{mean}(\min(\tau, |I(x',y')|)^\alpha))^{\frac{1}{\alpha}}} \quad (14)$$

$$I(x,y) \leftarrow \tau \tanh\left(\frac{I(x,y)}{\tau}\right) \quad (15)$$

4 Results

For testing the preprocessing chains frontal face images (fa and fb) of the FERET (Facial Recognition Technology) face database [10] are used. All preprocessing approaches were tested against 500 static runs. One run consists of 20 Reject-Tests

and about 4 Accept-Tests, all randomly chosen. The same images are used to test all methods, therefore they are called "static runs", this ensures meaningful comparability. The False Acceptance Rate (FAR) and False Rejection Rate (FRR) were calculated over all 500 runs to form the recognition rate of Table 1. One run consists of Enrollment and Evaluation/Test Phase, see Figure 1. In the Enrollment Phase templates of 3 face images are calculated and stored for later authentication. In the Evaluation/Test Phase one template of a face image is created and compared to the formerly stored Enrollment templates. By using a manually set threshold a face gets accepted or rejected. Wrong decisions are recorded in the FAR and FRR. Every processed image is preprocessed with the previously discussed approaches.

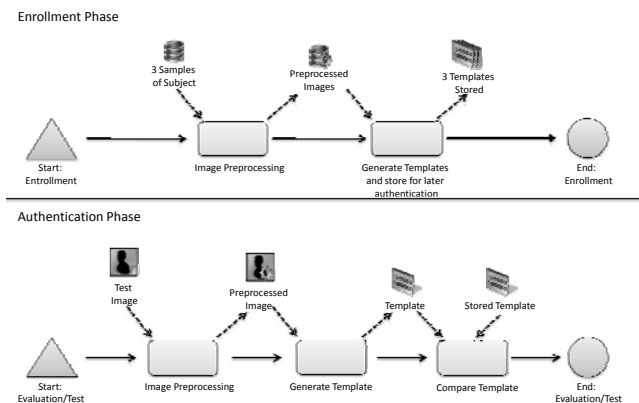


Fig. 1 General structure of the verification process

Table 1 Experimental results using different Illumination Compensation Methods

Illumination Compensation Method	Recognition Rate
No Method	90.8
Histogram Equalization	91.4
Daub4	89.5
Multi-Scale Retinex	90.7
TT	91.4

5 Conclusion

The major contribution of this paper was to answer the question if additional pre-processing steps can substantially improve the recognition rate of the advanced face recognition approach ELGBPHS. Face recognition papers based on the pattern description method LBP are usually using preprocessing steps before the actual face

recognition approach is applied. A major drawback of additional processing steps is longer execution time and higher memory usage. We wanted to examine if this extra cost is worth the effort.

Our evaluation shows that TT and Histogram Equalization achieve the highest recognition rate but summarized all tested methods show comparable results. Using an illumination compensation method as additional preprocessing step only causes an insignificant improvement. Our experiments indicate that ELGBPHS is a very robust face recognition approach. ELGBPHS itself compensates illumination variations and other disturbing conditions by means of the powerful texture description LBP.

We state to consider to trade-off if implementing and executing the mentioned preprocessing steps. In our opinion the robust ELGBPHS does not need preprocessing except Face Detection and Resize which should always be performed. This saves execution time and memory because the image data is considerably reduced.

References

- [1] Acharya, T., Ray, A.K.: *Image Processing: Principles and Applications*. Wiley-Interscience, Hoboken (2005)
- [2] Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J.(G.) (eds.) *ECCV 2004*. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
- [3] Bolme, D.: *Elastic bunch graph matching*. Master's thesis, Colorado State University (2003)
- [4] Du, S., Ward, R.: Wavelet-based illumination normalization for face recognition. In: *Proceedings of the 2005 IEEE International Conference on Image Processing (ICIP 2005)*, pp. 954–957. IEEE Computer Society, Los Alamitos (2005)
- [5] Gross, R., Baker, S., Matthews, I., Kanade, T.: *Face Recognition Across Pose and Illumination*. In: *Handbook of Face Recognition*, pp. 193–216. Springer, Heidelberg (2005)
- [6] Intel. *Opencv library* (2010), <http://opencv.willowgarage.com/wiki/>
- [7] Jobson, D., Rahman, Z., Woodell, G.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* 6(7), 965–976 (1997)
- [8] Kela, N., Rattani, A., Gupta, P.: Illumination invariant elastic bunch graph matching for efficient face recognition. In: *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW 2006)*, p. 42. IEEE Computer Society, Los Alamitos (2006)
- [9] Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* 42(3), 300–311 (1993)
- [10] Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The feret database and evaluation procedure for face recognition algorithms. *Image Vision Comput.* 16(5), 295–306 (1998)
- [11] Rahman, Z., Jobson, D.J., Woodell, G.A.: Retinex processing for automatic image enhancement. *J. of Electron. Imaging* 13(1), 100–110 (2004)

- [12] Serrano, A., de Diego, I.M., Conde, C., Cabello, E.: Recent advances in face biometrics with gabor wavelets: A review. *Pattern Recogn. Lett.* 31(5), 372–381 (2010)
- [13] Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE* 94, 1948–1962 (2006)
- [14] Su, Y., Shan, S., Chen, X., Gao, W.: Hierarchical ensemble of global and local classifiers for face recognition. *IEEE Trans. Image Process.* 18(8), 1885–1896 (2009)
- [15] Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* 19(6), 1635–1650 (2010)
- [16] Viola, P., Jones, M.: Robust real-time face detection. *Int. J. Comput. Vision* 57(2), 137–154 (2004)
- [17] Wiskott, L., Fellous, J.M., Krüger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal.* 19(7), 775–779 (1997)
- [18] Xie, S., Shan, S., Chen, X., Chen, J.: Fusing local patterns of gabor magnitude and phase for face recognition. *IEEE Trans. Image Process.* 19(5), 1349–1361 (2010)
- [19] Zhang, B., Shan, S., Chen, X., Gao, W.: Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition. *IEEE Trans. Image Process.* 16(1), 57–68 (2007)
- [20] Zhang, W., Shan, S., Qing, L., Chen, X., Gao, W.: Are gabor phases really useless for face recognition? *Pattern Anal. Appl.* 12(3), 301–307 (2009)
- [21] Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. *ACM Comput. Surv.* 35(4), 399–458 (2003)

Efficient Stamps Classification by Means of Point Distance Histogram and Discrete Cosine Transform

Paweł Forczmański and Dariusz Frejlichowski

Abstract. The problem of stamp recognition addressed here involves a multi-stage approach which includes stamp detection, localization and segmentation, features extraction and finally, classification. In this paper we focus on the two last stages, namely features extraction by means of Point Distance Histogram and Discrete Cosine Transform, and classification employing distance calculation by means of Euclidean metrics. The first stage which leads to automatic stamps segmentation has been described in several previous papers and it is based mainly on color segmentation. The feature extractor described here works on binary images of stamps and employs polar representation of points gathered in a histogram form, which is later reduced by means of Discrete Cosine Transform. At the classification stage, compact descriptors of stamps are compared according to the distance to the reference objects (class' centers), and the closest class is taken as the answer. The paper includes some results of selected experiments on real documents having different types of stamps. A comparison with the classical two-dimensional DCT calculated over the images is also provided to prove high discriminative power of the developed approach.

1 Introduction

The problem of stamps classification has emerged from the issue of seal imprint identification on bank checks, envelopes, and transaction receipts addressed in the scientific literature in the last 30 years, which has still not been solved in a satisfactory way [8, 11, 12]. There are few scientific works related to the problem of stamps recognition except several articles related to stamps detection [8, 11]. There are two possible reasons for this fact. One of them is the observation that many

Paweł Forczmański · Dariusz Frejlichowski

West Pomeranian University of Technology, Szczecin,

Faculty of Computer Science and Information Technology, Żołnierska Str. 52, 71–210
Szczecin, Poland

e-mail: [pforczmanski,dfrejlichowski}@wi.zut.edu.pl](mailto:{pforczmanski,dfrejlichowski}@wi.zut.edu.pl)

official documents are converted into digital domain and the analysis of paper-based documents is no longer needed. However, it should be remembered that there are still numerous documents which are gathered in libraries and archives and they also may be an objective of research and computer recognition of those documents can be a very interesting topic of investigations. On the other hand, there are many situations where stamping is still employed, i.e. in the field of official documents, identity cards, formal letters, certificates, travel identification documents, post office, etc. All these documents are being issued by formal authorities and will not change in the close future. Hence the stamping as the process of authentication will be present in our surrounding for next years and it is the main support of the works presented in this paper. This paper extends the research described in [4] in the area of descriptor calculation and further classification.

2 Algorithm Overview

The whole procedure of stamp recognition is divided into two main stages: stamp extraction and stamp classification. The first stage of initial processing has been presented in [4]. It leads from stamp detection to the stamp localization and extraction. Below we describe the second stage — stamp classification, which is divided into several steps: calculation of Point Distance Histogram (PDH), reduction of features dimensionality by means of Discrete Cosine transform (DCT) and distance calculation by means of Euclidean Metrics. The flow of computations is presented in Fig. 1. Detailed description of each step is presented in the following sections.

3 Features Extraction/Reduction

3.1 Stamps Characteristics

There are few features that can be employed to describe stamps found in the traditional paper-based documents. They come from the technical characteristics of the process of stamping and can be observed in most of documents throughout the centuries. In this paper we focus on official stamps as they play a very important role in practical tasks, as it was mentioned in the introduction. Typical stamp, no matter of the shape, consists of regularly-shaped silhouette with clearly visible text and possible ornaments. The geometrical features of stamps (e.g. shape, complexity, typical patterns) have evolved into *de facto* standards. The analysis of the problem shows that the shape of stamp defines its category, hence official stamps are in most cases round. There are 4 more classes of stamp shapes that are less popular: oval, rectangular, square and triangular. Sample members of these shapes are presented in Tab. 1. The color information is also very important since original documents include color stamps. The black ink is less frequently used for official stamping, and in many cases suggests a copy (not original document). However in the procedure described in this paper, we assume to have monochromatic images of stamps.

Table 1 Sample official stamps, often regular and without distinct decorations, divided into five classes.

class	exemplary stamps				
rectangular					
square					
triangular					
round					
oval					

3.2 Point Distance Histogram Representation

The localized stamps, when extracted from an image (the process was described in [4]), can be represented with a shape descriptor. There are lots of algorithms developed for the task of shape representation so far. In general, they can be divided into two main groups, according to the applied version of a shape — the whole region or the outline only. The exhaustive survey of the shape description algorithms is beyond the scope of this paper. Nevertheless, one can be referred to a review by Zhang and Lu [10].

In works presented in this paper the Point Distance Histogram (PDH) was applied. This selection comes from the advantages of this algorithm; they will be described later in this subsection. This shape representation algorithm was firstly presented in [1]. It draws benefits from two very popular in image analysis and efficient techniques — histogram and polar transform. Such combination is not new. For example the Distance Histogram [6] calculates the distances from the centroid to all pixels within the object. The achieved values are normalized according to the highest one. The Polar Histogram [7] is very similar, but it enhances the contribution of the transformation from Cartesian to polar coordinates. Despite of the similarity

between PDH, DH and PH there is an important difference. The DH and PH calculate the descriptions using all pixels belonging to a shape, including its interior. In the case of the PDH only the contour is taken into consideration. Hence, this method is much faster.

The most important advantages of the Point Distance Histogram are:

1. invariance to the problem of the starting point selection and direction of tracing the contour during the work with particular points,
2. invariance to scaling, achieved thanks to the normalization of the final histogram,
3. invariance to rotation, coming from the usage of the histogram,
4. invariance to translation within the image plane, thanks to the calculation of distances from particular points on the contour to the assumed point within the shape (e.g. centroid),
5. possibility of the usage of some other points than centroid as an origin of the transform, what can be useful in particular applications (e.g. in the presence of occlusion),
6. controlling of the behavior of the achieved descriptions on the basis on the parameter r , denoting the number of bins in the histogram.

The Point Distance Histogram was until now applied to several problems employing the analysis and recognition of shapes. It was successfully used in the recognition of erythrocyte shapes extracted from the digital grayscale MGG images of human blood [5]. It was also applied to the problem of trademark recognition [2]. Lately the advantages of the PDH descriptor were utilized in the general shape analysis [3]. This problem has properties similar both to the recognition and retrieval of shapes. However, the processed object usually does not belong to any of the template classes. The idea is that it is only similar to some basic shapes, e.g. triangle, circle, square. Thanks to this one can conclude how triangular, round, square, etc. an analyzed shape is. The general shape analysis can be applied to various tasks, where this method of coarse classification is desirable.

As it was mentioned, the Point Distance Histogram can use various methods for the derivation of the origin of the transform. In the simplest case the centroid is applied [1]:

$$O = (x_o, y_o) = \left(\frac{1}{N} \sum_{i=1}^N x_i, \frac{1}{N} \sum_{i=1}^N y_i \right), \quad (1)$$

where:

N — number of points in the contour,

x_i, y_i — a point in the contour, $i = 1, \dots, N$. Later two vectors are calculated — Θ for angles and P for radii [1]:

$$\rho_i = \sqrt{(x_i - x_o)^2 + (y_i - y_o)^2}, \quad \theta_i = \text{atan} \left(\frac{y_i - y_o}{x_i - x_o} \right). \quad (2)$$

The values in Θ are converted into nearest integers [1]:

$$\theta_i = \begin{cases} \lfloor \theta_i \rfloor, & \text{if } \theta_i - \lfloor \theta_i \rfloor < 0.5 \\ \lceil \theta_i \rceil, & \text{if } \theta_i - \lfloor \theta_i \rfloor \geq 0.5 \end{cases} . \quad (3)$$

Next, the elements in the vectors Θ and P are rearranged according to the increasing values in Θ . Resultant vectors are denoted as $\hat{\Theta}$ and \hat{P} . For equal values in $\hat{\Theta}$ only single element with the highest corresponding value in \hat{P} is left and the other ones are omitted. This way we eliminate repeating angular values by selecting only the one with the highest distance from the origin of the transform. As a result, only the outer points of the contour are selected, hence the method is less sensitive to the internal structure of a stamp. In effect we get two vectors with at most 360 elements, one for each integer angle (in degrees). Only the new vector \tilde{P} is required for further work, where $k = 1, 2, \dots, m; m \leq 360$. The elements in \tilde{P} are normalized according to the maximal one [11]:

$$M = \max_k \{\tilde{\rho}_k\}, \quad \tilde{\rho}_k = \frac{\tilde{\rho}_k}{M}. \quad (4)$$

Finally, the histogram is built — the elements in \tilde{P} are assigned to r bins in histogram, $\tilde{\rho}_k$ to h_k [11]:

$$h_k = \begin{cases} r, & \text{if } \tilde{\rho}_k = 1 \\ \lfloor r\tilde{\rho}_k \rfloor, & \text{if } \tilde{\rho}_k \neq 1 \end{cases} . \quad (5)$$

In the experiments described earlier in this subsection the parameter r was equal to 25 which have given satisfactory results. This value was pre-established experimentally. In the following section we describe the approach to reduce this value to obtain a much more compact representation.

3.3 Features Dimensionality Reduction by Means of DCT

One of the most frequently and widely used spectral transformations in the field of pattern recognition, but also in general image processing tasks, is the Discrete Fourier Transform (DFT) and its variant — Discrete Cosine Transform (DCT) [9]. By using the DFT coefficients calculated for the stamp image, it is possible to obtain features, which are invariant to periodic shifting in the image plane. We assume, however, that the PDH descriptor is invariant to the affine transforms, as it was shown above, therefore we use DCT as the method of obtaining spectral components. The use of DCT is motivated by the fact that in many simple practical applications, its basis functions are often used instead of the eigenfunctions from the Karhunen-Loève Transform (KLT). The other motivation is the widespread of fast DCT calculation algorithm in the field of digital image processing. On the other hand, it does make it possible to describe the stamp represented as its PDH descriptor sufficiently with a small feature set — the spectral DCT components. Like in the KLT, the high energy components tend to be in the beginning of the spectrum vector. If the initial PDH descriptor is represented by a vector H of size $1 \times r$

elements, then the calculation of spectral features of the DCT can be described in the following discrete form [9]:

$$S(u) = \sqrt{\frac{2}{r}} C(u) \sum_{k=0}^{r-1} H(k) \cos \frac{(2k+1)u\pi}{2r}, u = 0, \dots, r-1, \quad (6)$$

$$\text{where: } C(u) = \begin{cases} 2^{-1/2} & \text{for } u = 0 \\ 1 & \text{otherwise} \end{cases}.$$

The whole processing flow is presented below, in Fig. 1. As it can be seen, it involves simple vector manipulations, leading from PDH calculation to reduced (truncated) DCT spectrum. Finally, the shape representation is reduced to a few DCT coefficients which can be later compared using simple distance metrics (Euclidean metrics in this case).



Fig. 1 Process of feature extraction and reduction using PDH/DCT approach.

As a comparison, the typical flow of computation for two-dimensional DCT is presented in Fig. 2. As it can be seen, the procedure requires two-dimensional transformation (DCT2) and 'zig-zag' projection from two-dimensional representation into one-dimensional vector, which is later truncated. The final shape representation is also reduced to a few DCT coefficients, yet different from the first approach (Fig. 1), also possible to compare by simple distance metrics.

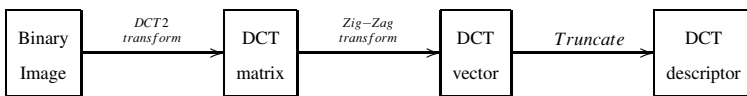


Fig. 2 Process of feature extraction and reduction using DCT2 approach.

Both above presented flows of computations were investigated in the context of recognition performance. The results are presented in the section 4.1.

4 Classification Procedure

In the experiments we employed distance metrics based on Euclidean distance as a measure of dissimilarity, since it is easy to compute and its interpretation is clear. The input of the classification stage is a test object (extracted stamp) which is processed using above described procedure, leading to the PDH/DCT descriptor, being

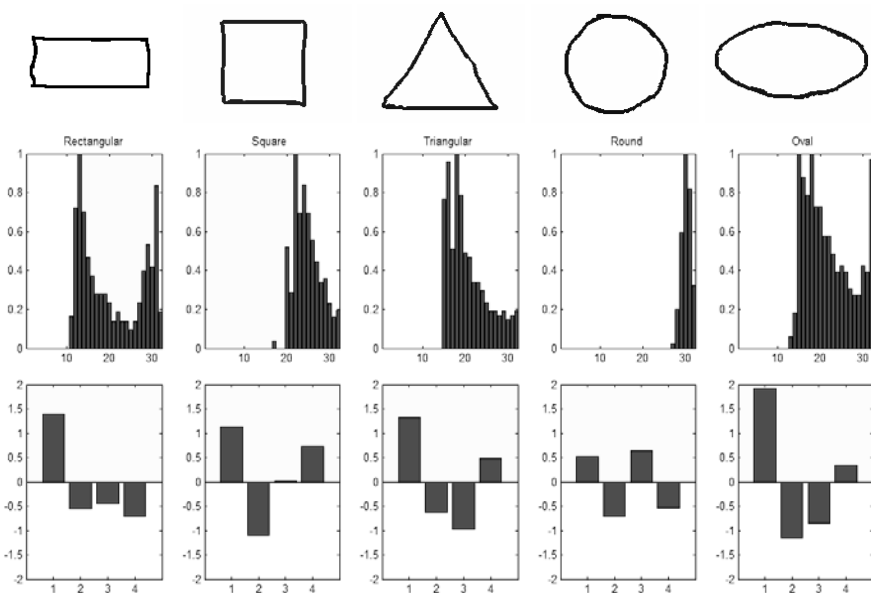


Fig. 3 Templates used in the process of PDH Descriptor calculation (upper row) and graphical representations of PDH (middle row) and their DCT-reduced representations (lower row).

a vector of very low-dimensionality. This vector is compared with J templates (descriptors) related to the most common observable stamp classes, giving a set of distance values. The lowest distance implies the class of the test stamp. In the Fig. 3 one can observe the templates (hand drawn) in the upper row, their PDH descriptors calculated for 32 bins (middle row) and first four components of DCT spectra of respective PDH descriptors (lower row).

As it can be seen from the above picture, the reduced vectors obtained by means of DCT transformation of PDH descriptors are unique for all classes, which suggests good discriminative power and efficient recognition.

4.1 Experiments and Results

In order to prove the assumed high discriminative power of PDH/DCT descriptor we performed the experiments on a database containing bitmaps of scanned stamps, extracted from freely available images found in the Internet. The initial preprocessing was performed using an approach described in [4]. All 140 stamps stored in binary matrices were divided into $J = 5$ classes: rectangular (48 elements), square (12 elements), triangular (11 elements), round (52 elements), and oval (17 elements). The classification procedure was described in the above section, however there were three variants of templates: ideal silhouettes drawn in the image editor (rectangle,

square, triangle, circle, and oval), hand drawn shapes (shown in Fig. 3) and average PDH/DCT descriptors calculated over all elements in each class in the database (as the comparison *niveau* only). The results of the experiments are presented in Tab. 2 where column 'ideal' represents ideal templates, 'hand drawn' — manually drawn images, and 'average' stands for mean descriptors calculated for all images in each class. It can be clearly seen that in most cases the best results give average descriptors calculated for all objects in the testing database. However it is a base for comparison only. The second best result is for hand drawn templates in a configuration 8 bins for PDH and 4 components after DCT.

Table 2 Recognition performance of PDH/DCT for different templates and different descriptor sizes

DCT comp.	ideal								hand drawn								average							
	PDH size (bins)								PDH size (bins)								PDH size (bins)							
	8	16	24	32	64	128	256		8	16	24	32	64	128	256		8	16	24	32	64	128	256	
1	38	21	23	22	22	24	19	49	51	53	46	41	45	51	61	58	58	53	54	55	49			
2	44	24	28	27	27	28	21	56	59	56	55	51	54	52	64	54	58	54	54	53	44			
4	71	67	66	54	54	56	54	75	73	74	73	74	74	74	76	74	75	74	74	74	72			
8	64	63	62	52	52	56	52	72	70	69	68	69	72	71	76	76	76	75	75	75	74			
16	-	59	53	47	47	51	49	-	66	67	66	69	69	64	-	76	76	76	76	75	76			
24	-	-	54	50	50	49	47	-	-	64	66	69	68	65	-	-	76	76	77	76	75			
32	-	-	-	51	51	49	48	-	-	-	66	70	68	63	-	-	-	76	77	76	75			
64	-	-	-	-	57	52	48	-	-	-	-	68	69	61	-	-	-	-	76	77	76			
128	-	-	-	-	-	59	53	-	-	-	-	-	67	61	-	-	-	-	-	77	76			
256	-	-	-	-	-	-	50	-	-	-	-	-	-	59	-	-	-	-	-	-	76			

The additional experiments were performed to prove high efficiency of the proposed method in a comparison to typical two-dimensional DCT performed on whole images. We employed the scheme depicted in Fig. 2. The results are presented in Tab. 3.

Table 3 Recognition performance of DCT2 for different templates and different spectrum size (number of components after 'zig-zag' scanning)

	DCT2 spectrum size									
	1	2	4	8	16	32	64	128	256	512
ideal	9	9	16	16	16	15	14	14	14	18
hand drawn	12	12	10	18	25	24	17	24	21	20
average	30	30	30	32	39	40	44	51	54	57

In order to obtain one-dimensional descriptors (vectors) we employed 'zig-zag' scanning of elements gathered in DCT2 matrices and selected first, most important components. As it can be seen from Tab. 3, the performance of such an approach is much lower. The only close results are for the case when we use 512 components of DCT spectrum, which is much larger than in the developed algorithm.

5 Summary

The main inspiration of the research presented in this paper was a shortage of algorithms that are able to classify stamps extracted from digital images as visual objects. The application area of this kind of a system is broad, ranging from law-enforcement forces, law offices, official archives and any other institutions that utilize stamping technique. Developed algorithm for stamp classification is divided into two stages. At the first stage, the processed stamp (binary matrix) is represented using a shape descriptor. In the paper the Point Distance Histogram was applied for this purpose. It was motivated by its advantages in this particular problem (e.g. invariance to the affine transforms of planar shape, generalization property). The second stage involves dimensionality reduction by means of Discrete Cosine Transform. Finally, the PDH/DCT descriptor is compared with a templates stored in the database using Euclidean metrics.

There are two main conclusions coming from the performed experiments. The first one is related to the PDH calculation, namely if the number of bins in the obtained histogram is too large, the recognition result are worse. It comes from the fact that in case of stamp images, which are small in general, the number of points on the silhouette can be also limited. Moreover, better results were achieved for hand drawn templates than for the ideal ones. It seems to be surprising on the first sight. However, it is obvious, when we see that the hand drawn templates have more points on the contour than the ideal ones. Despite the fact the PDH algorithm considers only the furthest points from the origin of the polar transform, there are more of them in the second case. Thanks to this more values are taken into consideration during deriving the histogram.

The second main conclusion is related to the application of Discrete Cosine Transform. The dimensionality reduction performed on descriptors calculated by means of PDH have shown, that only four first DCT components are useful for classification purposes. Increasing the number of DCT components is not justified, since it does not increase the overall recognition accuracy.

Finally, we presented classification employing classical two-dimensional DCT calculated over the image matrix which proved higher discriminative power of the developed approach.

The possible area of applications of this algorithm is broad and was briefly described in the introductory part of the paper. In comparison to the standard, direct application of DCT on images, developed descriptor gives much better results together with much lower dimensionality.

References

- [1] Frejlichowski, D.: The Point Distance Histogram for Analysis of Erythrocyte Shapes. *Polish Journal of Environmental Studies* 16(5b), 261–264 (2007)
- [2] Frejlichowski, D.: Influence of the Number of Bins on the Recognition Results Using Point Distance Histogram. *Electronics — Constructions, Technologies, Applications* (11), 58–60 (2009)

- [3] Frejlichowski, D.: An Experimental Comparison of Seven Shape Descriptors in the General Shape Analysis Problem. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010. LNCS, vol. 6111, pp. 294–305. Springer, Heidelberg (2010)
- [4] Frejlichowski, D., Forczmański, P.: General Shape Analysis Applied to Stamps Retrieval from Scanned Documents. In: Dicheva, D., Dochev, D. (eds.) AIMSA 2010. LNCS (LNAI), vol. 6304, pp. 251–260. Springer, Heidelberg (2010)
- [5] Frejlichowski, D.: Pre-processing, Extraction and Recognition of Binary Erythrocyte Shapes for Computer-Assisted Diagnosis Based on MGG Image. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010. LNCS, vol. 6374, pp. 368–375. Springer, Heidelberg (2010)
- [6] Saykol, E., Gudukbay, U., Ulusoy, O.: A histogram-based approach for object-based query-by-shape-and-color in image and video databases. *Image and Vision Computing* 23, 1170–1180 (2005)
- [7] Suau, P.: Robust artificial landmark recognition using polar histograms. In: Bento, C., Cardoso, A., Dias, G. (eds.) EPIA 2005. LNCS (LNAI), vol. 3808, pp. 455–461. Springer, Heidelberg (2005)
- [8] Ueda, K., Nakamura, Y.: Automatic verification of seal impression patterns. In: Proc. 7th. Int. Conf. on Pattern Recognition, pp. 1019–1021 (1984)
- [9] Watson, A.B.: Image Compression Using the Discrete Cosine Transform. *The Mathematica Journal* 4(1) (1994)
- [10] Zhang, D., Lu, G.: Review of shape representation and description techniques. *Pattern Recognition* 37, 1–19 (2004)
- [11] Zhu, G., Jaeger, S., Doermann, D.: A robust stamp detection framework on degraded documents. In: Proceedings - SPIE The International Society For Optical Engineering, vol. 6067 (2006)
- [12] Zhu, G., Doermann, D.: Automatic Document Logo Detection. In: The 9th International Conference on Document Analysis and Recognition (ICDAR 2007), pp. 864–868 (2007)

Color Image Retrieval Based on Mixture Approximation and Color Region Matching

Maria Luszczkiewicz-Piatek and Bogdan Smolka

Abstract. The paper introduces the extension of the color image retrieval method based on the approximation of the perceptual parameters. The proposed solution enables effective search for similar images regardless of the applied compression scheme not only taking into account the color palette and the presence of regions of the homogenous color within the image, but also their spatial arrangement. The proposed method utilizes the *Gaussian Mixture Modeling* combined with the *Bilateral Filtering* approach along with color matching method based on dominant region color. The evaluated results show that satisfactory retrieval results can be obtained regardless to applied compression schemes, preserving the spatial arrangement of the color regions in evaluated results.

1 Introduction

The image retrieval field is widely explored these days due to the enormous influence of the Internet technologies on every-day life. Therefore it is essential that an user is provided with powerful tools enabling efficient image retrieval regardless of image file format available in e.g. World Wide Web. The most important tools and methods are presented in [2, 9]. Since CBIR methods rely on feature extraction from images, color, because of its robustness to image deterioration, changes in size, resolution or orientation, is the most widely explored image feature in retrieval solutions. Moreover, color spaces, known to reflect human perception (e.g. $L^*a^*b^*$) widely applied in this field, are very suitable for applications such as proposed in

Maria Luszczkiewicz-Piatek

University of Lodz, Department of Applied Computer Science,
Banacha 22 Str, 98-238 Lodz, Poland

e-mail: mluszczkiewicz@math.uni.lodz.pl

Bogdan Smolka

Silesian University of Technology, Department of Automatic Control,
Akademicka 16 Str, 44-100 Gliwice, Poland

e-mail: bogdan.smolka@polsl.pl

this paper. However, very often search results are expected to share not only the color palette of the query image, but also its content, understood as the visual similarity of the query and the retrieved results. The presented method addresses both of the problems. Firstly, the retrieval success is not bound to the image compression rate, because of the applied approximation of the color histogram using Gaussian Mixture Model (GMM) counteracting the loss of color information in contrary to Vector Quantization approach do not addressing these problems, [6, 7, 8, 16]. Secondly, the image content similarity is reflected by the homogenous color regions, and more importantly, their spatial location.

Presented spatial color region matching is a tool for retrieval of images sharing spatial color composition of the query, resulting in closer similarity of evaluated results, understood as the information conveyed by scene depicted by analyzed images. Thus, the proposed solution can be useful in retrieving images of the same content but differently arranged within the image as a result of the changes in depicted scene between capture moments. The proposed method utilizes set of colors which represents image only by palette of 10 elements. Let us note that such an approach reflects the main image content, omitting unimportant details. Although, the methods based on histogram approach to image color lose spatial information of image, the application of proposed color region matching can be an answer to this disadvantage. In general, the approach to color region matching is based on division of the analyzed image into the number of blocks specifying the representative color of each block.

The proposed technique comprises of the following steps: firstly each color image within the database is analyzed in order to evaluate its representation by Gaussian Mixture. Let us note that for the incorporation of the information about the color distribution in the image, the color histogram, subjected to mixture modeling, is constructed using the approach inspired by the Bilateral Filtering (BF), [5]. On that basis the retrieval is evaluated producing the ranked list of results. As the similarity measure of the mixture models of the query and candidate images the *Earth Mover's Distance* (EMD), [15] is used. The EMD is based on the assumption that one of the histograms reflects "hills" and the second represents "holes" in the ground of a histogram. The measured distance is defined as a minimum amount of work needed to transform one histogram into the other using a "soil" of the first histogram. As this method operates on signatures and their weights using the approach based on the *Gaussian Mixture Models* we assigned as signature values the *mean* of each component and for the *signature weight* the weighting coefficient of each Gaussian in the model. The EMD computations were evaluated using the Y. Rubner algorithm implementation. This step concludes the first stage of the proposed method.

The second part of the algorithm introduces the information about the spatial arrangement of the colors within the image. The representative color is chosen for each of $n \times m$ image regions. The retrieval process, based on matching of region colors, produces the set of ranked results, which combined with ranking evaluated in the first part of the algorithm, provides the set of resulting candidate images.

The presented experiments were evaluated on the database of Webmuseum¹ which hosts over 1300 color images of paintings that span a wide range of artistic styles (see Fig. 1).

The presented paper is organized as follows. In Section 2 the image representation using mixture approach is briefly described. The proposed approach is also introduced. In Section 3 color region matching, being a complementary to mixture modeling is presented. The experimental results are shown and discussed in Section 4 along with drawn conclusions presented in Section 5.



Fig. 1 The exemplary images of the Webmuseum database. The collection comprises masterpieces representing wide range of artistic styles.

2 Image Representation by Mixture Modeling

The solution presented in this paper utilizes the $L^*a^*b^*$ space, [14]. Let us note that the L^* component is not taken into account. The first step in applying the proposed methodology is to construct the histogram $H(x,y)$ in the $a-b$ chromaticity space defined as $H(x,y) = N^{-1} \# \{a_{i,j} = x, b_{i,j} = y\}$, where $H(x,y)$ denotes a specified bin of a two-dimensional histogram with a - component equal to x and b - component equal to y , the symbol $\#$ denotes the number of elements in a bin and N is the number of image pixels.

Let us note that the application of many compression techniques can change and in consequence often cause a distortions and loss of the color information conveyed by the image. The proposed methodology can successfully deal with this undesirable property and assure the effective retrieval. Moreover, the drawback of various kinds of standard color histograms is that they do not take into account the spatial image structure. As a result images of the same color proportions but of different color arrangements will have the same chromaticity histogram. Our goal is to take into account the spatial arrangement of the image pixels. The first step of the incorporation of the region homogeneity information into the color histogram, is association

¹ www.ibiblio.org/wm/

of the weighting coefficients provided by the *Bilateral Filter* to each pixel in order to reflect their similarity to neighboring ones.

Weighting coefficients calculated for every image pixel are then taken into account during the constructing of chromaticity histogram. Therefore, each pixel of the original image is incorporated into $a - b$ histogram with the weight reflecting its color similarity to neighboring pixels. In details, weights are computed according to the following scheme, [11]:

$$w_{x,y} = \frac{1}{n} \sum_{(i,j) \in W} \exp\left(-\frac{\|c_{x,y} - c_{i,j}\|}{h}\right)^{k_1} \exp\left(-\frac{d_{i,j}}{\delta}\right)^{k_2} \quad (1)$$

where $c_{i,j}$ and $c_{x,y}$ denote the color pixels at positions (i, j) and (x, y) respectively, h is the color difference smoothing parameter, $d_{i,j}$ is the topological distance between the pixel at position (i, j) and the filtering window W centered at (x, y) and δ is a spatial smoothing parameter equal to the diameter of the square filtering window. The pixel number n in W was set to be equal to 10% of N and we assumed $k_1 = k_2$. The retrieval based on GMM+ BF sorting produces rank list of candidate images: $r(BF)_\varphi$, where φ is the index of the analyzed image.

Let us note that such an approach does not determine the spatial layout of the regions of homogenous color within the image because it takes into account only two aspects of pixel similarity: topological closeness and color similarity.

The next step of the indexing method, is the construction of a $a - b$ histogram approximation using the GMM and the Expectation-Maximization algorithm [4], which provides a set of model parameters serving as an image signature that can be used for retrieval purposes [10, 11]. Figure 4 illustrates the modeling process on the $a - b$ plane.

The important feature of the GMM is that it is able to restore the color information lost during the image compression (see Fig. 4 b and c). Therefore this part of proposed solution provides means for uncorrupted selection of a set of candidate images which are the most similar to a given query on the basis of their color content. However, the problem of the spatial arrangement of the comprised colors within the image is still not addressed. The next Section will discuss this problem in details.

3 Color Region Matching

In order to achieve the satisfactory retrieval results not only the color palette should be taken into account but also the spatial arrangement of the regions of the homogenous color. Therefore, the proposed technique utilizes the BF approach along with color region matching. If only the overall color palette of the image is taken into account, the spatial organization of the images is not analyzed. Let us note that the BF approach to spatial arrangements of the colors comprised in the image, does not recognize the differences like that between images shown in green frame in Fig. 3(a), because only the image palette and the fact that homogenous regions are present are analyzed. Thus, these two images share the color palette and chromaticity

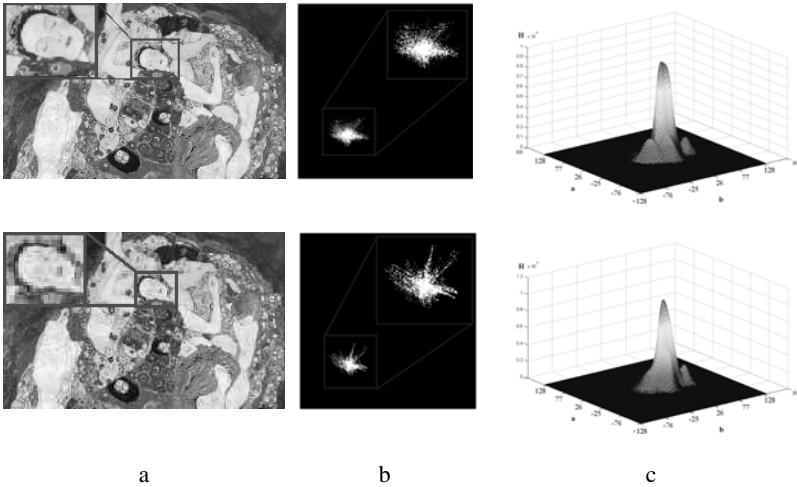


Fig. 2 The visualization of the approximation of the chromaticity histogram (b) using Gaussian Mixture Model (c) of *Gustav Klimt's "Virgin"* (a) for the original image (upper row) and lossy compressed (10% of the original file size using JPEG method.) and resized (10% of original image size) version (bottom row).

histogram, but the spatial arrangement of image colors reflecting conveyed information is not analyzed. Therefore, in order to reflect the similarity to given query image, the spatial color structure is taken into account during the retrieval process. Moreover, ranking of the retrieved candidate images is based on the match of the representative colors of the corresponding image regions.

In order to facilitate the choice of the representative color of every analyzed region, each color image is firstly subjected to color segmentation using the Hill Climbing Method, [1, 13]. Thus, color image is represented as 3 dimensional histogram in $L^*a^*b^*$ color space. Than application of the Hill Climbing technique provides the initial color clusters locations for the *k-means* method evaluating color segmentation. The approach presented in this work assumes that each color present in the image can be associated to one of $k = 10$ clusters (colors). In result, the set of triples (L^*, a^*, b^*) is chosen.

Then, each analyzed image is divided into $R_1 \times R_2$ regions and the dominant color $r_{i,j}$ of each region where $i \in 1, \dots, R_1$ and $j \in 1, \dots, R_2$ is assigned. For purpose of this work $R_1 = R_2 = 5$ was assumed. The number of the color regions is a key factor influencing the match effectiveness. The more regions are chosen, the more accurate the color structure of the image is reflected. However, the influence of possible color distortions caused by pixels of outlier colors to their neighborhood is also increased. The dominant color is determined on the basis of the mean value of all related pixels. In order to determine the similarity among the corresponding regions of the candidate image and the query image the sum of the Euclidean distance among regions' representative color is evaluated.

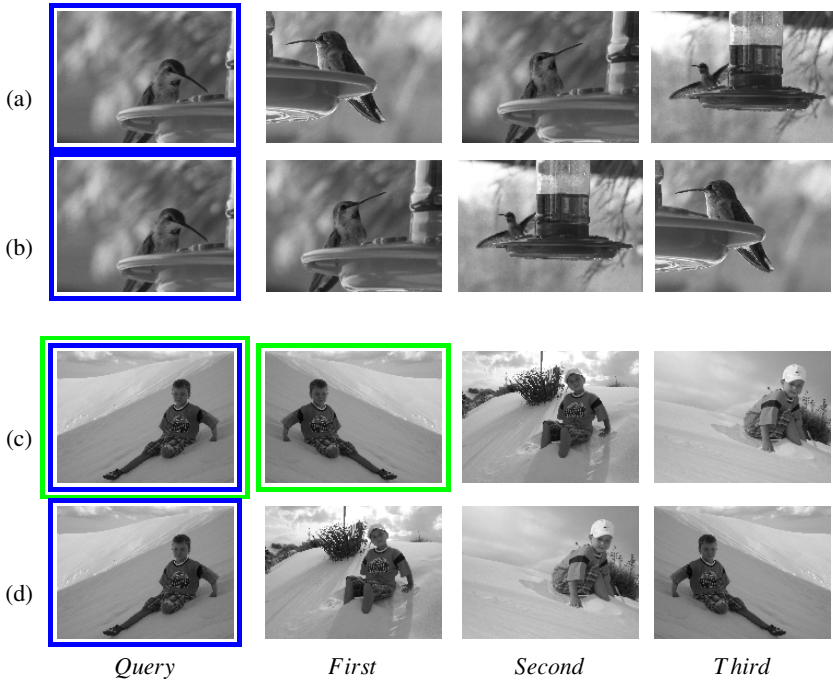


Fig. 3 Upper rows (*a* and *c*) show the retrieval results evaluated on the basis of the GMM of color histogram combined with the BF approach. The additional application of color region matching enables to take into account the image content reflected by spatial organization of image colors, (*b* and *d*). Query images are placed in blue frames. Green-framed images (*c*) share the chromaticity histogram, but spatial arrangement of color content is different.

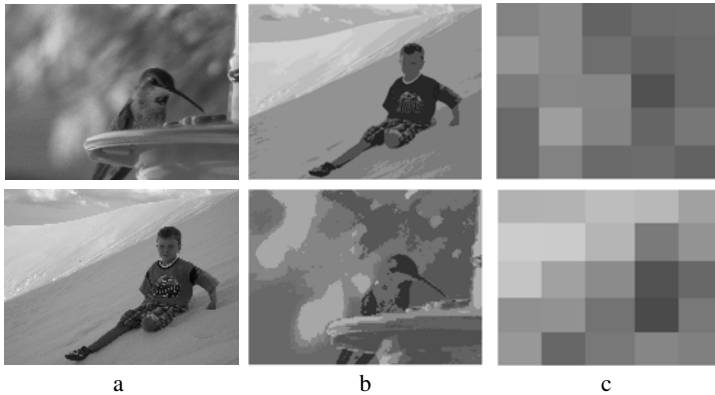


Fig. 4 The process of choosing the representative color of image (a) regions. Image is firstly segmented into $k = 10$ groups (b). For each of $R_1 \times R_2 = 25$ image region the representative color is chosen (c).

Then, the color images of the analyzed database are ranked according to their spatial color arrangement match to a given query.

The final color matching rank z_{image} between candidate images and the query image is evaluated in following manner: two sets of ranked candidate images are taken into account as: $z_{image} = sortA(z_{BF}(im) + z_{CM}(im))$ where $sortA$ denotes sorting operation in ascending order, z_{BF} and z_{CM} denote the ranks related to *GMM + BF* and *Color Match* approaches.

4 Retrieval Results

The application of the proposed method enables to retrieve candidate images which not only share the color palette of the original query images, regardlessly of the applied compression scheme, but also present regions of the homogenous color placed in the same location as in the original image. Figure 5 illustrates the proposed solution applied on database of paintings contained in the WebMuseum.

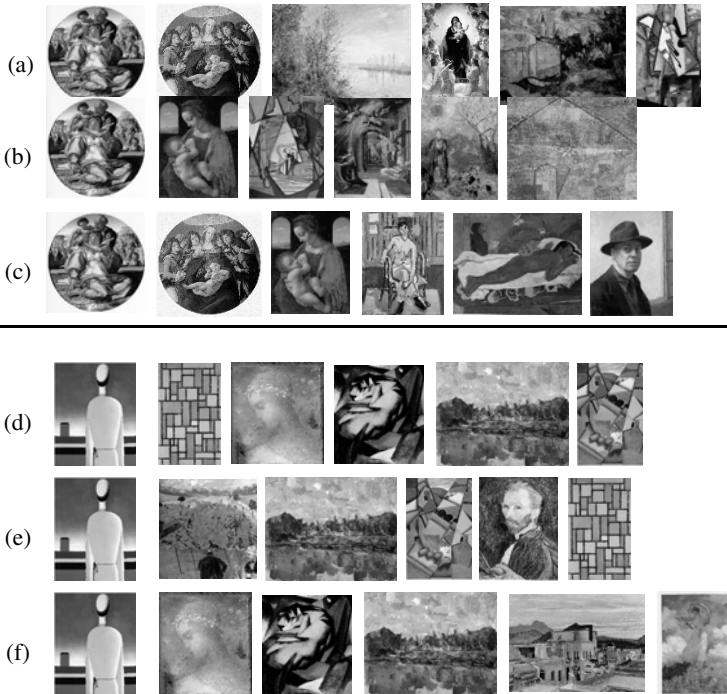


Fig. 5 The comparison of the retrieval results performed on the database of Webmuseum obtained for the various retrieval methods based on color image representation using: GMM with no additional weighting (a), GMM combined with Bilateral Filtering method (b), and methods (a) and (b) combined with color region matching (c). The results were evaluated for images transformed to *JPEG₂₅* (25% size of original image), and represented in the *CIE_{Lab}* color space.

Let us note that the images comprising matching regions of similar representative color are promoted over those sharing only the color palette of the query. Regions are identified rather by their representative hue than by actual color because of possible color distortions introduced by lossy coding or possible noise. Combining these two retrieval schemes enables effective image search, as shown in Fig. 5. It can be noticed that candidate images share a spatial layout of the query, comprising color regions of matching colors in corresponding locations, despite of slight differences in color palettes of the query and retrieved images. However, these differences do not disturb perceptual similarities between query and retrieved images, which are more visible than those between query and candidate images sharing only overall color palette. This phenomenon can be noticed as retrieved candidate images preserve color region composition of the query, which can be seen in Fig. 5. Moreover, the application of the proposed methodology provides more accurate retrieval results as shown in Fig. 5f, than when spatial image structure is not analyzed (d and e). This example clearly proves that when no spatial information about color location within the image is available, the retrieved images (e.g. Fig. 5d first retrieval result) could have no meaningful relation to the query although its obvious relation in color content reflected by chromaticity histogram.

5 Conclusions

The retrieval methods based only on the approximation of the color histogram by mixture model or combined with bilateral approach provide correct results only when the overall image palette is taken into account. However, these methods only partially preserve color region information, i.e. the homogeneous color areas are identified by application of bilateral filter approach, but their spatial arrangement within the image is not taken into account during the retrieval process. The proposed solution² combines the ranking evaluated on the basis of the image palette content and that based on representative region color. Therefore, this scheme is robust to color distortion due to image manipulation and can assure correct retrieval results for images of various (even large) compression rates (resulting in sometimes severe loss of color information) because of reconstruction abilities of mixture approach related to approximation of color histogram (as seen in Fig. 4), [10, 11]. Moreover, bilateral approach enables to retrieve images not only sharing the color palette of query but also promoting images of similar spatial arrangement of colors in form of homogeneous color regions. However, it is crucial for any content search to associate corresponding color regions present on query and retrieved images. Thus, it is reasonable to incorporate an approach, which ranks images according to their spatial color arrangement match to a given query. These two steps (i.e. GMM+BF and color region matching) are evaluated simultaneously and independently in the proposed solution, providing robustness to change of pixel colors due to lossy coding (because of the application of *GMM+BF*). Moreover, a coarse color region matching is able to refine retrieval results evaluated in first part of algorithm.

² For color version of this paper visit www.marialu.w.interii.pl

Acknowledgments. This work has been supported by the Polish Ministry of Science and Higher Education under R&D grant no. N N516 374736 from the Science Budget 2009-2011.

References

- [1] Achanta, R., Estrada, F., Wils, P., Sausstrun, S.: Salient region detection and segmentation. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) ICVS 2008. LNCS, vol. 5008, pp. 66–75. Springer, Heidelberg (2008)
- [2] Antani, S., Kasturi, R., Jain, R.: A survey of the use of pattern recognition method for abstraction, indexing and retrieval. *Pattern Recognition* 1, 945–965 (2002)
- [3] Bilmes, J.: A Gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Markov Models. Technical Report, University of Berkeley, ICSI-TR-97-021 (1997)
- [4] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistics Society* 39, 1–38 (1977)
- [5] Elad, M.: On the origin of the bilateral filter and ways to improve it. *IEEE Transactions on Image Processing* 11(10), 1141–1151 (2002)
- [6] Gray, R.M.: Gauss mixture vector quantization. In: *Proceedings of IEEE ICASSP*, vol. 3, pp. 1769–1772 (2005)
- [7] Jeong, S., Won, C. -S., Gray, R.M.: Image retrieval using color histograms generated by Gauss mixture vector quantization. *Computer Vision and Image Understanding* 94(1-3), 44–66 (2004)
- [8] Jeong, S.: Distributional distances in color image retrieval with GMVQ-generated histograms. LNCS, vol. 3569, pp. 465–475. Springer, Heidelberg (2005)
- [9] Liu, Y., Zhang, D., Lu, D., Ma, W.Y.: A survey of content-based image retrieval with high level semantics. *Pattern Recognition* 40, 262–282 (2007)
- [10] Luszczkiewicz, M., Smolka, B.: Gaussian Mixture Model based retrieval technique for lossy compressed color images. In: Kamel, M.S., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 662–673. Springer, Heidelberg (2007)
- [11] Luszczkiewicz, M., Smolka, B.: Application of bilateral filtering and Gaussian Mixture modeling for the retrieval of paintings. In: *Proceedings of the International Conference on image Processing (ICIP 2009)*, pp. 77–80 (2009)
- [12] McLachlan, G., Peel, D.: *Finite Mixtures Models*. John Wiley & Sons, Chichester (2000)
- [13] Ohashi, T., Aghbari, Z., Makinouchi, A.: Hill-climbing algorithm for efficient color-based image segmentation. In: *Proc. of IAESTED International Conference on Signal Processing, Pattern Recognition, and Applications* (2003)
- [14] Plataniotis, K.N., Venetsanopoulos, A.N.: *Color Image Processing and Application*. Springer, Berlin (2000)
- [15] Rubner, Y., Tomasi, C., Guibas, L.J.: The Earth Mover’s Distance as a metric for image retrieval. *International Journal of Computer Vision* 40(2), 99–121 (2000)
- [16] Vasconcelos, N.: Minimum probability of error image retrieval. *IEEE Transactions on Signal Processing* 52(8), 2322–2336 (2004)

LAMDA Methodology Applied to Image Vector Quantization

E. Guzmán, J.G. Zambrano, I. García, and Oleksiy Pogrebyak

Abstract. In this paper we present a novel approach to image vector quantization (VQ) based on Learning Algorithm for Multivariate Data Analysis (LAMDA methodology). The proposed algorithm, named VQ-LAMDA, employs a codebook generated by LBG algorithm, which must be normalized to obtain a new codebook representation, named LAMDA-codebook. The classification phase of the LAMDA methodology uses the fuzzy binomial distribution to determine marginal adequacy degrees between input vectors and LAMDA-codebook. Then, to obtain global adequacy degrees, are used the operators min-max and product and a linear convex function. Finally, using the Maximum Adequacy rule are calculate a set of indices of the codewords, to which every input vector belongs. The computer simulation results demonstrate that the proposed algorithm provides better performance than other VQ methods based on the LGB algorithm, at the peak signal-to-noise ratio (PSNR) parameter.

1 Introduction

Vector Quantization (VQ) is a fundamental element of applications involving data compression. VQ is the most used scheme due to the fact that it results in lower distortion for a given rate, in comparison with scalar quantization schemes. On the other hand, however, the complexity and memory requirements of a vector quantizer increases exponentially with the number of samples in a vector (vector dimension) [8]. To solve this problem, have been proposed algorithms based on neural networks, associative memories and fuzzy logic.

E. Guzmán · J.G. Zambrano · I. García
Universidad Tecnológica de la Mixteca
Carretera a Acatlima Km. 2.5, Huajuapán de León, Oaxaca, México
e-mail: eguzman@mixteco.utm.mx

Oleksiy Pogrebyak
Centro de Investigación en Computación del IPN
Av. Juan de Dios Bátiz, Cd. de México
e-mail: olek@pollux.cic.ipn.mx

An unsupervised learning model, called the Self-Organizing Map (SOM) was proposed by Teuvo Kohonen as a learning algorithm for neural networks of competitive learning. The SOM have also been used with a great deal of success in creating new VQ schemes [15], [16], [14].

Amerijckx C. *et al.* they applied the SOM at both quantification and codification stages of the image compressor. At the quantification stage, the SOM algorithm creates a correspondence between the input space of stimuli, and the output space constituted of the codebook elements (codewords, or neurons) derived using the Euclidean distance. At the entropy coder stage, a differential entropy coder uses the topology-preserving property of the SOMs resulted from the learning process and the hypothesis that the consecutive blocks in the image are often similar [3].

Recently, Guzmán E. *et al.* proposed a new fast search algorithm for vector quantization based on Extended Associative Memories (FSA-EAM) [9], [10]. This algorithm uses the Extend Associative Memories (EAM) to create an EAM-codebook (associative network) applying the EAM training stage to the codebook produced by the LBG algorithm. The FSA-EAM VQ process is performed using the recalling stage of EAM. This process generates a set of the class indices to which each input vector belongs. The main advantage offered by this algorithm is high processing speed and low demand of resources (system memory), while the encoding quality remains competitive.

A VQ process has two big fields of investigation, the codebooks generation and the search process of the reconstruction vector that presents a best matching to an input vector. In this paper, we propose to apply to the search process a novel algorithm for VQ of images based on Learning Algorithm for Multivariate Data Analysis (LAMDA).

This paper is organized as follow. In next section, the operation of LBG algorithm is explained. In section 3, a brief theoretical background of LAMDA methodology is given. Then, in section 4 the algorithm proposed, VQ-LAMDA algorithm for image compression, based on Learning Algorithm for Multivariate Data Analysis, is presented. Numerical simulation results obtained are provided and discussed in Section 5; furthermore, this section includes a comparative analysis with traditional methods of VQ. Finally, the section 6 contains the conclusions of this paper.

2 LBG Algorithm

A generalized Lloyd clustering algorithm (GLA) proposed by Y. Linde, A. Buzo and R. Gray, called the LBG algorithm, is typically used to generate VQ codebooks [17]. This algorithm is built upon the hypothesis that the optimization of the codewords in each stage is based on the condition that the outcome in the previous stage is optimum, resulting in practically the best codebook (at least locally) [6]. The codebook generation algorithm consists of the following sequence of steps:

Step 1. Initialization. Start with an initial codebook, formed by a finite set of n -dimension vectors, denoted by $C = \{\mathbf{y}^i : i = 1, 2, \dots, N\}$, and a set of training vectors $X = \{\mathbf{x}^j : j = 1, 2, \dots, M\}$, $k = 0$ and select distortion threshold ϵ .

Step 2. Partitioning. Given $C^{(k)} = \{y^i : i = 1, 2, \dots, N\}$, find all training vectors x^j that are closer to y^i than to any other y^m , in other words, find the minimum distortion partition $P(C^{(k)}) = \{S^i : i = 1, 2, \dots, N\}$ of the set of training vectors: $x^j \in S^i$ if $d(x^j, y^i) \leq d(x^j, y^m)$, for all $i \neq m$. Compute the average distortion (Euclidean distance).

Step 3. Codebook update. Iteration $k = k + 1$, in each partition $S^{i(k)}$, generate a new codeword by applying the centroid condition.

Step 4. Convergence Check. Stop when process converges, if $\frac{d^{k-1} - d^k}{d^k} \leq \epsilon$. Otherwise go to Step 2.

The Fig. 1 shows a VQ basic scheme based on the LBG algorithm (LBG-VQ). In this scheme, an image of $h \times w$ pixels is divided in M blocks of $d \times d$ dimensions; this blocks represent n -dimensional ($n = d \times d$) input vectors $X = \{x^j : j = 1, 2, \dots, M\}$. The LBG-VQ algorithm consists of an N -level and n -dimensional quantizer Q , that makes a mapping of a input vector in Euclidean space \mathcal{R}^n , into a finite subset C of \mathcal{R}^n , that containing N reconstruction vectors, called code vectors or codewords, $Q : \mathcal{R}^n \rightarrow C$, where $C = \{y^i : i = 1, 2, \dots, N\}$ and $y^i \in \mathcal{R}^n$ [7].

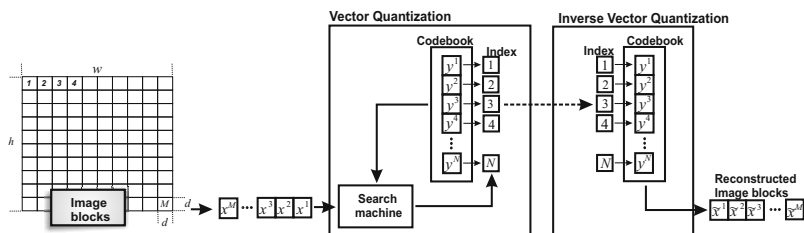


Fig. 1 LBG vector quantization scheme.

3 Theoretical Background of LAMDA Methodology

The Learning Algorithm for Multivariate Data Analysis (LAMDA) is an incremental conceptual clustering method based on fuzzy logic, which can be applied in the processes of formation and recognition of concepts (classes) [1], [19], [21]. LAMDA has the following features:

- The previous knowledge of the number of classes is not necessary.
- The descriptors can be qualitative, quantitative or a combination of both.
- Recognition of concepts is based on the maximum adequacy rule (MA).
- The α parameter serves to control the selectivity of the classification.

3.1 Operation of LAMDA Methodology

LAMDA methodology calculates a degree of adequacy, by means of a membership function between the descriptors of the input vectors (x^j) and the descriptors of

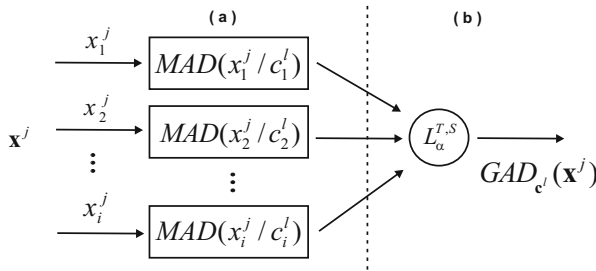


Fig. 2 LAMDA basic structure.

all previously established classes (\mathbf{c}^l). The set of objects can be described by $X = \{\mathbf{x}^j : j = 1, 2, \dots, M\}$ and any object can be represented by a vector $\mathbf{x}^j = (x_1, \dots, x_n)$. The set of classes can be described by $C = \{\mathbf{c}^l : l = 1, 2, \dots, N\}$ and any class can be represented by a vector $\mathbf{c}^l = (c_1, \dots, c_n)$ [2], [19], [21].

3.2 Marginal Adequacy Degree

Given an object \mathbf{x}^j and a class \mathbf{c}^l , LAMDA computes for every descriptor the marginal adequacy degree (MAD), between the value of component x_i of object \mathbf{x}^j and the value that the component c_i takes in \mathbf{c}^l (see figure 2a). Hence, a MAD vector can be associated with an object \mathbf{x}^j . This stage generates N MADs, and this process is repeated iteratively for every object with all classes [22]. The probability distributions that use LAMDA methodology are: Fuzzy Binomial, Fuzzy Binomial-Center, Fuzzy Binomial-Distance and Gaussian [2], [11]. The Fuzzy Binomial distribution was selected to be used to calculate the MADs in this implementation, due to that is the function that executes fewer operations, resulting in a low computational complexity.

3.3 Global Adequacy Degree

The Global Adequacy degree (GAD) is obtained by aggregating or summarizing of all marginal information previously calculated, given by the N MADs of an object \mathbf{x}^j relative to class \mathbf{c}^l , using T-norms and S-conorms and a linear convex T-S function ($L_{\alpha}^{T,S}$), where α is called exigency level (see figure 2b) [2], [19], [21].

LAMDA methodology uses the following T-norms and their dual S-conorm: min-max, Product, Lukasiewicz, Yager and Hammacher. In this implementation min-max and Product, were selected because they are the that execute fewer operations.

Finally, computed the GAD of the object \mathbf{x}^j related to all classes, and using the MA rule, \mathbf{x}^j will be placed in the highest degree adequacy class.

LAMDA has been applied to different domains, to name just a few: medical images [5], pattern recognition [20], detection and diagnosis of failures of

industrial processes [13], biological processes [4], distribution systems of electrical energy [18], processes for drinking water production [11], monitoring and diagnosis of industrial processes [12].

4 Proposed Algorithm: VQ-LAMDA

This section describes the proposed scheme, denominated VQ-LAMDA; for this purpose, we use the LAMDA methodology, furthermore, the codebook generated by LBG algorithm as the set of classes used in the encoding process.

4.1 LAMDA-Codebook Generation

The LAMDA codebook is computed in two phases.

Phase 1. LBG codebook generation. In this phase, a codebook based on LBG algorithm is generated. This phase is a supervised process; the training set used in the codebook generation is formed by image blocks. Let $\mathbf{x} = [x_i]$ be a n -dimensional vector, which represents an image block; then, the training set is defined by $\mathbf{A} = \{\mathbf{x}^j : j = 1, 2, \dots, M\}$. The result of this stage is a codebook denoted as $C = \{\mathbf{c}^l : l = 1, 2, \dots, N\}$, where $\mathbf{c} = [c_i]_n$.

Phase 2. LBG codebook normalization. Before using the LBG codebook, and to maintain coherence with fuzzy logic, it must be normalized through the use of (1).

$$c_i = \frac{\tilde{c}_i - c_{min}}{c_{max} - c_{min}} = \frac{\tilde{c}_i}{2^L - 1} \quad (1)$$

where, $i = 1, 2, \dots, n$, \tilde{c}_i is the descriptor before to be normalized, c_i is the normalized descriptor, $c_{min} = 0$ and $c_{max} = 2^L - 1$; in the context of image processing, L is the number of bits necessary to represent the value of a pixel. The limits (minimum and maximum) of the descriptors values are the limits of the data set.

The set of normalized descriptors is denominated **LAMDA-codebook** and is used in the **VQ-LAMDA** scheme.

4.2 VQ-LAMDA Scheme

The VQ scheme proposed performs the classification task according to membership criterion computed in four phases.

Phase 1. Image row vectors normalization: Before using the image descriptors in the VQ-LAMDA scheme, and to maintain coherence with fuzzy logic, they must be normalized through the use of (2).

$$x_i = \frac{\tilde{x}_i - x_{min}}{x_{max} - x_{min}} = \frac{\tilde{x}_i}{2^L - 1} \quad (2)$$

Table 1 T-Norms and S-Conorms.

Operator	T-Norm (Intersection)	S-Conorm (Union)
min-max	$\min(MAD(x_i^j/c_i^l))$	$\max(MAD(x_i^j/c_i^l))$
Product	$\prod_{i=1}^n MAD(x_i^j/c_i^l)$	$1 - (\prod_{i=1}^n MAD(x_i^j/c_i^l))$

where, $i = 1, 2, \dots, n$, \tilde{x}_i is the descriptor before to be normalized, x_i is the normalized descriptor, $0 \leq \tilde{x}_i \leq 1$, $x_{min} = 0$ and $x_{max} = 2^L - 1$, L is the number of bits necessary to represent the value of a pixel. The minimum and maximum values of the descriptors represent the limits of the data set.

Phase 2. Calculate the Marginal Adequacy Degree. MADs are calculated for each descriptor x_i^j of each input vector \mathbf{x}^j with each descriptor c_i^l of each class \mathbf{c}^l . For this purpose, we used the Fuzzy Binomial Distribution:

$$MAD(x_i^j/c_i^l) = (\rho_i^l)^{x_i^j} (1 - \rho_i^l)^{(1-x_i^j)} \tag{3}$$

where, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, M$ and $l = 1, 2, \dots, N$. For this distribution function $\rho_i^l = c_i^l$.

Phase 3. Calculate Global Adequacy Degree. This stage determines the grade of membership of each input vector \mathbf{x}^j to each class \mathbf{c}^l , by means of a linear convex function (using (4)) and the use of T-norms and S-conorms that are shown in Table 1.

$$L_\alpha^{T,S} = GAD_{c^l}(\mathbf{x}^j) = (\alpha) \cdot T(MAD(x_i^j/c_i^l)) + (1 - \alpha) \cdot S(MAD(x_i^j/c_i^l)) \tag{4}$$

where $\alpha \in [0, 1]$, $T \leq L_\alpha^{T,S} \leq S$, $T = L_1^{T,S}$ (intersection) and $S = L_0^{T,S}$ (union).

Phase 4. Obtaining the index. Finally, this phase generates the index of the class to which the input vector belongs. The index is determined by the GAD that presents the maximum value (MA rule).

$$index = \max(GAD_{c^1}(\mathbf{x}^j), GAD_{c^2}(\mathbf{x}^j), \dots, GAD_{c^l}(\mathbf{x}^j)) \tag{5}$$

5 Experimental Results

In this section, we present the experimental results obtained when the VQ-LAMDA is used in the process of VQ of images. First, we compared the VQ-LAMDA performance with respect to LBG and FSA-EAM VQ algorithms. Second, we analyzed the average number and type of operations used by the VQ-LAMDA, LBG and the FSA-EAM VQ algorithms [9]. For this purpose, the following set of test images of size 512×512 with 256 gray levels were used in simulations.

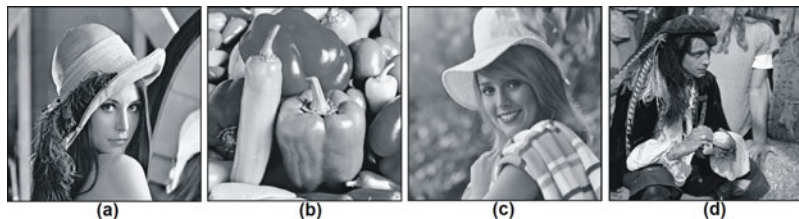


Fig. 3 Set of test images: (a)Lena, (b)Peppers, (c)Elaine, (d)Man.

In order to evaluate performance of the proposed algorithm, VQ-LAMDA, we used a popular objective performance criterion called peak signal-to-noise ratio (PSNR), which is defined as

$$PSNR = 10 \log_{10} \left(\frac{(2^L - 1)^2}{\frac{1}{M} \sum_{i=1}^M (p_i - \tilde{p}_i)^2} \right) \tag{6}$$

where L is the number of bits per pixel, M is the number of pixels in the image, p_i is the i -th pixel in the original image, and \tilde{p}_i is the i -th pixel in the reconstructed image.

Table 2 Summary of results of VQ-LAMDA with the codebook of different sizes.

Algorithm	Exigency level(α)	T-norms and S-conorms	Codebook size	Distortion measure (dB)			
				Elaine	Lena	Man	Peppers
VQ-LAMDA (Binomial)	0.5-0.6	Min-Max	64x16	27.29	27.17	25.16	26.27
	1	Product		28.92	28.80	26.81	27.80
	LBG	-		28.91	27.14	24.20	26.38
	FSA-EAM	-		28.31	26.31	23.28	25.12
VQ-LAMDA (Binomial)	0.5-0.6	Min-Max	128x16	27.43	27.63	25.60	26.75
	1	Product		29.69	29.73	27.53	28.60
	LBG	-		29.68	28.21	24.93	27.18
	FSA-EAM	-		29.07	27.47	23.96	26.06
VQ-LAMDA (Binomial)	0.5-0.6	Min-Max	256x16	27.45	27.72	25.85	26.88
	1	Product		30.31	30.44	28.25	29.13
	LBG	-		30.30	29.08	25.49	27.64
	FSA-EAM	-		29.68	28.39	24.55	26.54
VQ-LAMDA (Binomial)	0.5-0.6	Min-Max	512x16	27.53	27.89	25.89	27.15
	1	Product		30.76	31.05	28.73	29.74
	LBG	-		30.74	29.98	26.02	28.21
	FSA-EAM	-		30.11	29.25	25.09	27.04

Table 3 Operations required by LBG, FSA-EAM and VQ-LAMDA to quantify a pixel.

Algorithm	Codebook size	The average number of operations per pixel					
		CMP	±	×	SQRT	Div	Power
VQ-LAMDA <u>min-max</u>	128	264	512	384	-	128	256
(Binomial) <u>product</u>		8	256	256	-	128	256
LBG		8	387	129	8.062	-	-
FSA-EAM		136	128	-	-	-	-
VQ-LAMDA <u>min-max</u>	256	528	1024	768	-	256	512
(Binomial) <u>product</u>		16	512	512	-	256	512
LBG		16	771	257	16.062	-	-
FSA-EAM		272	256	-	-	-	-
VQ-LAMDA <u>min-max</u>	512	1056	2048	1536	-	512	1024
(Binomial) <u>product</u>		32	1024	1024	-	512	1024
LBG		32	1539	513	32.062	-	-
FSA-EAM		544	512	-	-	-	-

The first experiment has the objective to compare the image distortion generated by the LBG, FSA-EAM and the two variants of VQ-LAMDA algorithm. In Table 2, we presented the findings of this experiment, in the combination of the fuzzy binomial distribution with the T-norm and S-conorm product, we chose the exigency level equal to 1 (although the range of the exigency level, in the which were obtained the best results is 0.51 to 1.), so the linear convex function is reduced by half, and consequently the number of operations is reduced. On the other hand, in the combination of the fuzzy binomial distribution with the T-norm and S-conorm min-max, we can not reduce the operations due that the range of exigency level, in the which were obtained the best results is 0.5-0.6. Table 2, shows that the combination of the fuzzy binomial distribution with the T-norm and S-conorm product has better performance than the LBG and FSA-EAM algorithms in the PSRN parameter.

In the second experiment, we compare the complexity of the VQ-LAMDA LBG and FSA-EAM algorithms. Table 3, summarizes computation complexities in terms of the average number of operations executed per pixels and type of operations. This experiment was performed for codebook size 128, 256 and 512, which are the most popular codebook sizes. With regard to time complexity, Table 3, shows that the proposed algorithm is competitive with the LBG algorithm but it is overcome by the FSA-EAM algorithm.

6 Conclusions

In this paper, we have proposed the use of LAMDA methodology in a VQ algorithm for image compression. The use of LAMDA at the VQ stage of an image compressor has demonstrated a high competitiveness in its efficiency in comparison to VQ traditional methods based on LBG algorithm. Considering the average number of

operations executed per pixels, our proposed algorithm is competitive with the LBG algorithm but it is overcome by the FSA-EAM algorithm. On the other hand, based on objective analysis, where objective performance criterion PSRN was used, we can conclude that when our proposal uses the combination of the fuzzy binomial distribution with the T-norm and S-conorm product, the quality of the image processed provides improvement in comparison with the quality obtained by traditional VQ methods based on the LBG algorithm.

References

- [1] Aguilar-Martin, J., Estruch, R.S., Waissman, J.: Knowledge-based Signal Analysis and Case-based Condition Monitoring of a Machine Tool. In: Joint 9th IFSA World Congress and 20th NAFIPS International Conference Proceedings, vol. 1, pp. 286–291 (2001)
- [2] Aguilar-Martin, J., Agell, N., Sánchez, M., Prats, F.: Analysis of Tensions in a Population Based on the Adequacy Concept. In: Escrig, M.T., Toledo, F.J., Golobardes, E. (eds.) CCIA 2002. LNCS (LNAI), vol. 2504, pp. 17–28. Springer, Heidelberg (2002)
- [3] Amerijckx, C., Legat, J.-D., Verleysen, M.: Image Compression Using Self-Organizing Map. *Systems Analysis Modelling Simulation* 43(11), 1529–1543 (2003)
- [4] Jean-Charles, A., Doncescu, A., Aguilar-Martin J.: A Fuzzy Clustering Approach for Supervision of Biological Processes by Image Processing. In: EUSFLAT, European Society for Fuzzy Logic and Technology, pp. 1057–1063 (2005)
- [5] Chan, M., Aguilar-Martin, J., Piera, N., Celsis, P., Vergnes, J.: Classification techniques for feature extraction in low resolution to topographic evaluative images: Application to cerebral blood flow estimation. In: 12th Conf. GRESTI. Groupe d'Etudes du Traitement du Signal et des Images (1989)
- [6] Chang-Qian, C., Soo-Ngee, K., Sivaprakasapillai, P.: A modified generalized Lloyd algorithm for VQ codebook design. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1996, vol. 1, pp. 542–545 (1996)
- [7] Gersho, A., Gray, R.M.: Vector quantization and signal compression. Kluwer Academic Publishers, Massachusetts (1992)
- [8] Gray, R.M.: Vector Quantization. *IEEE ASSP Magazine* 1, 4–9 (1984)
- [9] Guzmán, E., Pogrebnyak, O., Sánchez, L., Yáñez-Márquez, C.: A fast search algorithm for vector quantization based on associative memories. In: Ruiz-Shulcloper, J., Kropatsch, W.G. (eds.) CIARP 2008. LNCS, vol. 5197, pp. 487–495. Springer, Heidelberg (2008)
- [10] Guzmán, E., Pogrebnyak, O., Yáñez, C., Manrique, P.: Vector Quantization Algorithm based on Associative Memories. In: Aguirre, A.H., Borja, R.M., Garcíá, C.A.R. (eds.) MICAI 2009. LNCS, vol. 5845, pp. 324–336. Springer, Heidelberg (2009)
- [11] Hernández, H.R.: Supervision et diagnostic des procédés de production d'eau potable. Ph.D. dissertation. l'Institut National des Sciences Appliquées de Toulouse, France (2006)
- [12] Isaza, C.V.: Diagnostic par Techniques d'Apprentissage Floues: Conception d'une Methode de Validation Et d'Optimisation des Partitions. Ph.D. dissertation. l'Université de Toulouse, France (2007)

- [13] Kempowsky, T.: Surveillance de Procédes a Base de Methodes de Classification: Conception d'un Outil d'aide Pour la Detection et le Diagnostic des Defaillances. Ph.D. dissertation. l'Institut National des Sciences Appliquées de Toulouse, France (2004)
- [14] Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg (2000)
- [15] Kohonen, T.: Automatic formation of topological maps of patterns in a self-organizing system. In: Oja, E., Simula, O. (eds.) Proc. 2SCIA, Scand. Conf. on Image Analysis, pp. 214–220 (1981)
- [16] Kohonen, T.: Self-organizing formation of topologically correct feature maps. Biological Cybernetics 43(1), 59–69 (1982)
- [17] Linde, Y., Buzo, A., Gray, R.: An Algorithm for Vector Quantizer Design. IEEE Transactions on Communications 28(1), 84–95 (1980)
- [18] Mora, J.J.: Localización de fallas en sistemas de distribución de energía eléctrica usando métodos basados en el modelo y métodos basados en el conocimiento. Ph.D. dissertation. Universitat de Girona, Catalua-España (2006)
- [19] Piera, N., Desroches, P., Aguilar-Martin, J.: LAMDA: An Incremental Conceptual Clustering Method. LAAS. Laboratoire d'Automatique et d'Analyse des Systems, Report No. 89420, pp. 1–21 (1989)
- [20] Piera-carreté, N., Desroches, P., Aguilar-Martin, J.: Variation points in pattern recognition. Pattern Recognition Letters 11, 519–524 (1990)
- [21] Piera, N., Aguilar-Martin, J.: Controlling Selectivity in Nonstandard Pattern Recognition Algorithms. IEEE Trans. on Systems, Man and Cybernetics 21(1), 71–82 (1991)
- [22] Waissman, J., Ben-Youssef, C., Vázquez, G.: Fuzzy Automata Identification Based on Knowledge Discovery in Datasets for Supervision of a WWT Process. In: 3rd Int. Conf. Sciences of Electronic Technologies of Information and Telecommunications. SETIT (2005)

Detection of Semantically Significant Image Elements Using Neural Networks

Jagoda Lazarek and Piotr S. Szczepaniak

Abstract. Detection of semantically significant image elements is a key task in the pattern recognition field. There are a lot of methods solving this problem. The main steps of the approach presented in this paper are colour transformation and detection, filtering and neural classification of contours.

1 Introduction

Detection of image elements is crucial for correct pattern recognition [1, 2] and it is the first step of the process automatic interpretation of the image content [3, 4, 5]. The key to solve the task of recognition of significant image elements is their detection on an image; it means that it is important to find the areas of the image, where such elements potentially exist. Preprocessing only a part of image gives the possibility to improve efficiency of image processing. The processing mentioned involves shortcomings elimination and finding significant details. There are a lot of filters which may be used. Filters are differentiated in terms of computational complexity, depending on the number of pixels. Because of this, the smaller the area, the more efficient the processing. This is why appropriate methods need to be chosen to make the searching area smaller. Moreover, the significant image descriptors are colour and shape. Further analysis determines correctness of the regions found. In this paper, we focus on detection of shapes having the form of polygons. The examples of practical importance are standard road signs.

The paper is organized as follows. First (section 2), the general description of the approach is presented. In sections 3, 4 and 5, the main steps of the approach are given, i.e. colour detection, filtering and final pattern recognition. The results of the practical examination of the method are given in section 6. The paper concludes with closing remarks.

Jagoda Lazarek · Piotr S. Szczepaniak

Institute of Information Technology Technical University of Lodz,
Wolczanska 215, 90-924 Lodz, Poland

e-mail: alma@stud.ics.p.lodz.pl

2 Approach

The proposed method is based on a priori known specific features of polygons (road signs in particular), such as shape and colour. Polygons, here related to groups of road signs, can be divided according to their features, e.g. number and length of sides (triangle, parallelogram, etc.).

The types of shapes used for road signs are related to their meaning; there are warning signs, prohibition signs, mandatory signs and informational signs. Each of them has a well defined shape; it may be a triangle, a square, a rhomb, an octagon or a circle. Also colours occurring on signs are exactly defined as red, yellow, blue and some others, but the first three are sufficient for identification of the type of the sign.

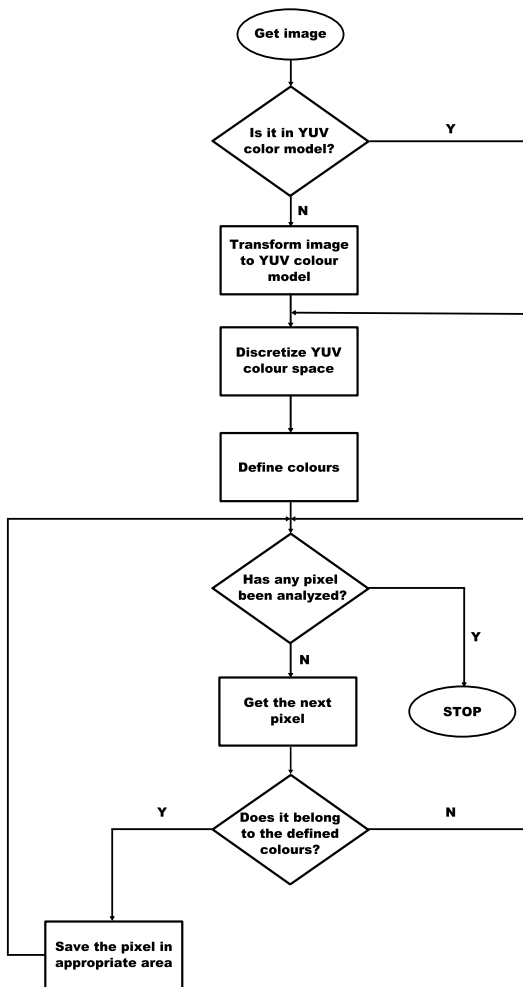


Fig. 1 The first part of the algorithm

The presented approach uses significant road signs descriptions mentioned earlier to select interesting parts of image being subject to further processing. To find the particular road sign on image's part, information included in contours is analysed. It is result of Sobel filtering. Finally, to assess a particular region on account of road sign's existing, neural network is used. In Fig. 1 and Fig. 2 the algorithm of the proposed method is presented.

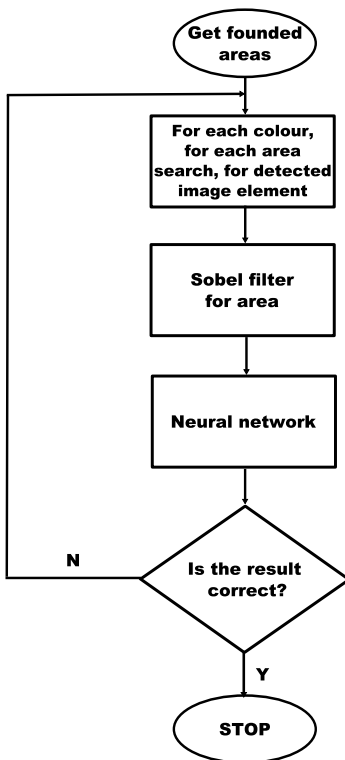


Fig. 2 The second part of the algorithm

3 Colour Segmentation

Colour segmentation is an important part of image processing, which gives the possibility to process only the selected parts of an image, determined by colours. Road signs have strictly predefined colours such as yellow, blue, red and black or white (pictograms) according to the sign type. Segmentation of the first three is sufficient to detect those image areas on which road signs may exist.

There real weather conditions can be different, e.g. rain, snow, or strong light. Consequently, the collection of images used for the examination of the method must involve a sufficient number of images having different qualities depending on weather conditions. It should be noticed that sometimes photographs may be

damaged because of too long or too short exposure, or blurring. Everything mentioned above has influence on the colour of the pixel. It follows that it is necessary to define the colours typical for road signs.

It is also necessary to choose the appropriate model of colour. RGB colour space is widely used in cameras. This is why photographs are represented by this model. But it is not the best choice, because in this model there is no possibility to distinguish information about colour and brightness. These problems can be avoided with colour space transformation. Conversion from RGB to YUV colour space gives the possibility to have separate information about brightness Y and colour U and V. Because of it, the YUV colour space has been chosen.

It is well known, that transformation of RGB into YUV (except for separation information of colour from brightness) is given by the following relation

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

After transformation, each pixel is checked and qualified to the correct class of colour. Only two logical expressions are needed to check the pixels. It is better than using RGB colour model which requires checking of six conditional instructions.

Assuming, for example, that there are defined arrays Y, U, V and 10 levels of discretization it is possible to determine if a pixel processed belongs to one of the predefined classes. Let us take a pixel $a=(y,u,v)$. Some computations are needed to check to which class pixel "a" belongs.

$$\begin{aligned} y' &= \lfloor (y/(25.5 + \alpha)) \rfloor \\ u' &= \lfloor (u + 111.18) / (22.236 + \alpha) \rfloor \\ v' &= \lfloor (v + 156.825) / (31.365 + \alpha) \rfloor, \text{ where } \alpha \ll 1 \end{aligned}$$

If $Y[y']$ AND $U[u']$ AND $V[v']$ is true than pixel "a" is in the class defined by Y, U and V (cf. [10, 11]).



Fig. 3 Segmentation of yellow colour from the image (see the square right on the top)

Definitions of colours are based on many samples from images. For the segmentation of red, yellow and blue, the right number of discretization levels is needed. For clarity, only segmentation of yellow colour is shown.

4 Filtering

Sobel filter is used for edge-detection. There are two 3×3 kernels applied separately to the image to find vertical and horizontal edges. Edge-detection filters base on evaluation of the gradient which gives information about brightness changing from one pixel to another. The final result is given by approximation of the gradient magnitude, on the basis of the first previously calculated derivatives [4, 6]. Kernels are defined as follows

$$Sobel_{horizon} : \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, Sobel_{vertical} : \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

To obtain the filtered image, some further computations are needed. Convolution masks are used to estimate gradient in vertical and horizontal directions

$$img_1 = img * Sobel_{horizon} \tag{1}$$

$$img_2 = img * Sobel_{vertical} \tag{2}$$

Result of Sobel filter is given by gradient magnitude

$$img_{Sobel} = \sqrt{img_1^2 + img_2^2} \tag{3}$$

which gives the image presented in Fig 4 contained important information needed in further processing.



Fig. 4 Contours of image’s part given by Sobel filter

5 Recognition

Image contours contain information of elements shape, which makes the recognition easy for a human. For computer recognition, one needs to apply computer-oriented methods. Here, artificial neural network (multilayer perceptron) was used to solve the problem of polygon detection. Note that this task, easy for a human, is not trivial for a computer, which acts in a different way.

The chosen architecture collects three layers (input, hidden and output), cf. Table 1. The hidden and output layers were equipped with the standard sigmoid activation function given by eq. (4) and the standard back-propagation method of learning was applied [7, 8, 9]. The momentum term has been used, and the values of parameters are given in Table 2. The course of the learning is shown in Fig. 5.

$$\varphi(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

Table 1 Architecture of the neural network

Layer	Neuron's number
Input layer	39
Hidden layer	45
Output layer	1

Table 2 Parameters

Parameter's name	Value
Learning rate	0.9
Momentum rate	0.6
Bias	1.0
Global error	0.001

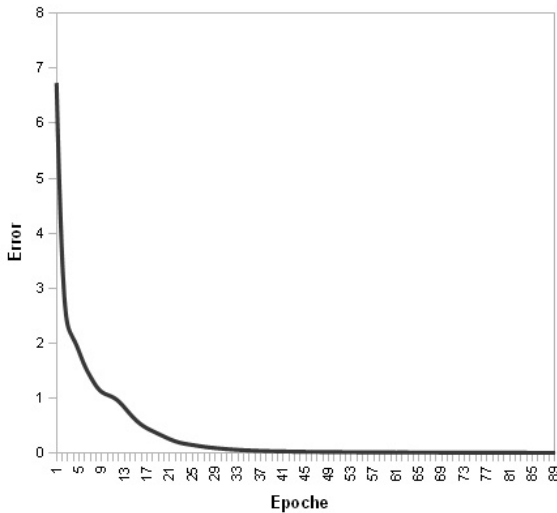


Fig. 5 Learning process

6 Tests and Concluding Remarks

Tests done on different types of images provide the possibility to observe satisfying effectiveness. The images analyzed here were taken in different weather conditions, such as snow or rain. Part of them was of poor quality because of too long or too short exposure, or due to blurring.

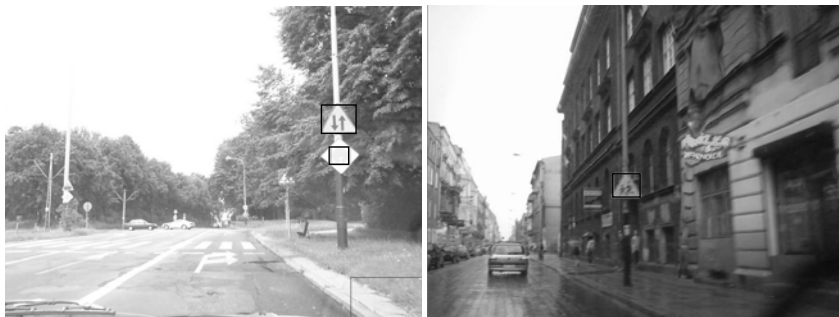


Fig. 6 Examples of images (strong light, blurred) and detection results (see contours in the form of squares)



Fig. 7 Examples of images (bad weather conditions, short exposure, rain) and detection results (see contours in the form of squares)

Tests have been done with the laptop with Intel Processor 2.0GHz, 1GB RAM; 96% of road signs (100 images, 640×480 px) have been detected correctly, in about 550 ms. The results show a significant potential of this approach and a possibility to use it in many fields, in which time of processing is essential.

To sum up, the method presented in this paper gives the possibility to find polygons on images taken even if in bad weather conditions or damaged by blurring, which is an important advantage because we aim at correctness of detection in any situation. The performed experiments confirm that polygons which are partially covered by other objects can also be found. The method is computationally effective, which is important when the time of detection becomes a significant user requirement. The use of the method for detection of other types of shapes as presented in the paper needs definition and numerical extraction of relevant figure features.

References

- [1] Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, Heidelberg (2006)
- [2] Tadeusiewicz, R., Flasiński, M.: *Rozpoznawanie obrazów (Image Recognition)*, Wydawnictwo Naukowe PWN (1991) (in Polish)
- [3] Tadeusiewicz, R., Ogiela, M.: *Medical Image Understanding Technology*. Springer, Heidelberg (2004)
- [4] Tadeusiewicz, R., Ogiela, M., Szczepaniak, P.S.: Notes on a Linguistic Description as the Basis for Automatic Image Understanding. *Int. Journal Appl. Math. Comput Sci.* 19(1), 143–150 (2009)
- [5] Szczepaniak, P.S., Tadeusiewicz, R.: The Role of Artificial Intelligence, Knowledge and Wisdom in Automatic Image Understanding. *Journal of Applied Computer Science JACS* 18(1), 75–85 (2010)
- [6] Watkins Ch., Sadun A., Marenka S.: *Nowoczesne metody przetwarzania obrazu*. Wydawnictwa Naukowo-Techniczne, Warszawa (1995) (in Polish)
- [7] Tadeusiewicz, R.: *Sieci neuronowe (Neural Networks)*. Akademicka Oficyna Wydawnicza RM, Warszawa (1993) (in Polish)
- [8] Haykin, S.: *Neural Networks. A Comprehensive Foundation*. Macmillan, New York (1994)
- [9] Rojas, R.: *Neural Networks. A Systematic Introduction*. Springer, Heidelberg (1996)
- [10] Malina, W., Smiatacz, M.: *Cyfrowe przetwarzanie obrazów*. Akademicka Oficyna Wydawnicza EXIT, Warszawa (2008) (in Polish)
- [11] Bruce, J., Balch, T., Veloso, M.: Fast and Cheap Color Image Segmentation for Interactive Robots. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000), Workshop on Interactive Robotics and Entertainment*, Pittsburgh, vol. 3, pp. 2061–2066 (2000)

The Effect of Multiple Training Sequences on HMM Classification of Motion Capture Gesture Data

Michał Romaszewski and Przemysław Głomb

Abstract. Hidden Markov Models (HMM) have been successfully applied to classification of various types of data, including human gestures. However, finding an optimal size of the training set for HMMs trained with multiple sequences and choosing a set of parameters for high performance is not a trivial task. We would like to address those issues by presenting results obtained using classifier based on HMM and Vector Quantisation applied to the set of a human gesture recordings. We use HMM as a model of a single gesture, and assess its recognition performance for multiple data sequences consisting of repetitions of selected gestures, performed by different persons with varying speed of movement. Additionally we intend to verify a reference database of 22 gestures for use in future experiments.

Keywords: gesture recognition, gesture interface, Vector Quantisation, HMM, gesture database.

1 Introduction

The use of gestures in Human-computer interface (HCI) promises to make many types of HCI interactions more effective and natural. However, the construction of gesture-based interface is not easy. Typical problems include the difficulty in clearly defining an adequate set of gestures and large differences between the various repetitions of a defined gesture.

Efficient gesture recognition is, and probably will be, a key element of a gesture-based HCI system. The result of capturing the motion of a human body can be efficiently stored in the form of data vectors corresponding to the bending of joints, and the speed of movement of individual body parts (e.g hands, fingers). HMMs are a natural tool to use for processing this kind of data and were successfully applied in multiple publications i.e. [1], [3].

Michał Romaszewski · Przemysław Głomb
Institute of Theoretical and Applied Informatics of PAS
Bałtycka 5, 44-100 Gliwice, Poland
e-mail: micchal,przemg@iitis.pl

While the performance of classifiers based on HMM and Vector Quantisation (VQ) is unquestionable, it is at the same time strongly affected by several issues like the size of available data set, the proper choice of HMM parameters, and initial setup. To address those issues, this paper contains the results concerning the impact of the training set size on the performance of HMM-based detector, trained with multiple gesture sequences. Additionally the analysis concerning the choice of the initial HMM parameters is provided.

Our research is focused on a human motion data, but the exact definition of a gesture is not obvious (for example in [11] the motion of a whole human body is considered while in [3] the gesture is represented by a hand motion only). To answer the need for a comparative set of a well-defined natural gestures the first instance of a gesture database (GDB) was prepared containing the data of the 22 natural gestures performed by different people at different speed rates. An additional goal of this paper is to verify the current instance of the database and draw conclusions about its further development.

1.1 Related Work

An excellent introduction to HMM usage is provided in [9] where it is noted that the result of the reestimation is a local maximum of the likelihood function, and whether it equals to global maximum depends on the initial values. [9] notes that good initial estimates are either helpful or essential (depending on the HMM type), and suggests iterative algorithm of data-dependent segmentation based on k-means clustering verified by reestimation process.

[10] provides an analysis and the extension of the content presented in [9] including the incorporation of scaling coefficients in formulas for multi-sequence identification.

In-depth analysis of issues related to the recognition of gestures may be found in [12] where authors used the HMM with VQ for the recognition performed on the set of nine different mouse gestures. Authors achieved high detection rate using bakis-topology discrete HMMs for both isolated and continuous recognition.

In [11] HMMs are used to detect gestures captured using video camera. Authors analyzed both hands and whole body movement and provided an interesting insights about extracting gesture features.

[3] presents the application of HMM for recognition of isolated and continuous gestures for Arabic numbers, captured using a video camera. Authors provided results for different number of HMM states and three selected HMM topologies.

[7] presents the gesture recognition with HMM used for creation of a Tangible User Interfaces. To capture gesture data the authors used a custom capturing devices called Smart Micrel Cubes, equipped with accelerometers. The authors also provided an in-depth analysis of the computational complexity related to the implementation of HMM based solutions on micro-controllers.

2 Method

The objective of the approach is to test the performance of a gesture data classifier based on VQ and HMM. The HMMs used in the experiments are trained with multiple sets of both raw and preprocessed data and initialised using different starting parameters (i.e. HMM topology).

2.1 The Recognition System

A single motion capture recording produces a time series sequence of vectors $\mathbf{X} = (\mathbf{x}_t)$, where each $\mathbf{x}_t \in \mathbb{R}^l$ represents data from l sensors (finger bend, hand rotation etc) captured in time moment t . We map the vectors \mathbf{x}_t to one of the m symbols s_t using Vector Quantisation (VQ):

$$Q: \mathbb{R}^l \rightarrow \mathbb{Z}_m \quad \mathbf{x}_t \mapsto s_t. \quad (1)$$

We use a well known Linde-Buzo-Gray VQ algorithm [5] with Euclidean distance.

To model the symbol sequences $\mathbf{S} = (s_t)$ we use a Hidden Markov Model λ with n states and a symbol set of cardinality m , defined by a starting state probability vector Π , a state-to-state transmission probability matrix \mathbf{A} and a symbol emission probability matrix \mathbf{B} :

$$\lambda = (\Pi, \mathbf{A}, \mathbf{B}), \quad \Pi = [\pi_i]_{n \times 1}, \quad \mathbf{A} = [a_{ij}]_{n \times n}, \quad \mathbf{B} = [b_{ij}]_{n \times m}. \quad (2)$$

Given a set of HMMs $\{\lambda_1, \lambda_2, \dots, \lambda_k\}$ and a sequence \mathbf{S} we can use forward algorithm [9] to compute the probability $P(\mathbf{S}|\lambda_i)$ and classify \mathbf{S} with $\operatorname{argmax}_{i=1\dots k} P(\mathbf{S}|\lambda_i)$. In practice, due to limited machine precision, we use the logarithm of probability $-\log P(\mathbf{S}|\lambda_i)$.

Our main focus in experiments is the detection rate γ , which we define as follows. We are given k gestures, each modelled by a separate HMM. A motion sequence \mathbf{S} known to be of gesture i is *recognised*, if $\operatorname{argmax}_{j=1\dots k} P(\mathbf{S}|\lambda_j) = i$. The γ is computed by dividing the number of recognised sequences n_r by number of all recognition attempts n_a done during a test, thus $\gamma = \frac{n_r}{n_a} \in \langle 0, 1 \rangle$.

2.2 Initialization and Data Preprocessing

As noted in [9], initialisation is an important step in building an effective classifier, as the performance of HMM is highly dependent on its initial conditions. A standard approach for model identification is choosing an initial model λ_0 , then using iterative Baum-Welch algorithm to optimize matrices for correspondence with modeled sequence \mathbf{S} .

To establish the effect of topology (configuration of a matrix \mathbf{A}) on the detection rate, we enumerate the following commonly used topologies (figure 1): ‘Ergodic’, where the transition from each state a_{ij} to every other state is possible; ‘Bakis’ and

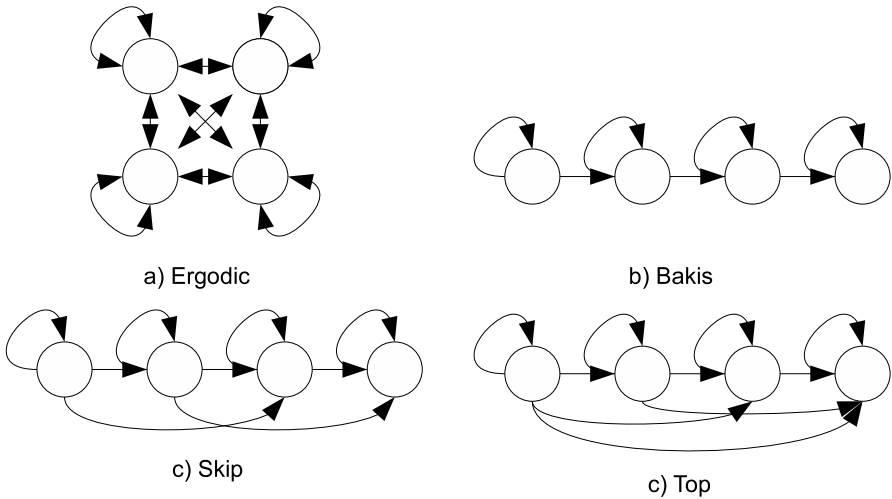


Fig. 1 Hidden Markov Model topologies used in the experiments.

‘Skip’ models, with the possibility of transition or skipping of the next state and the ‘Top’ model where only backward transitions are forbidden. Since in the last state of the topologies (with exception of the Ergodic model) only the recurrent transition are possible, we’ve included the last three configurations in two versions: ‘with absorbing state’ (regular, recurrent only) and ‘without absorbing state’ (also transition from the last state to the first state is possible).

We use uniform initialization for matrix \mathbf{A} (equal probability of jumping to the states allowed by the topology), random for matrix \mathbf{B} , and for either $\Pi = \begin{bmatrix} 1 \\ n \end{bmatrix}$ (Ergodic) or $\Pi = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T$ (all other topologies).

We focus on multiple training sequences for gesture model identification. As the formulas in [9] for this case do not consider scaling coefficients, which are necessary in practical applications (the probabilities quickly diminish below machine precision), we use the multi-sequence identification formulas provided in [10].

To test the impact of the data preprocessing on the detection quality, additional two sets were created from the initial raw gesture data set. In the first set, the data was resampled using linear interpolation, effectively normalising the time domain. The number of resampled data vectors for every gesture was set to $n = 100$. In the second set, aside from interpolation, the sensor readings were a subject to z-score normalization:

$$[\mathbf{x}_t]'_i = \frac{[\mathbf{x}_t]_i - \mu_i}{\sigma_i} \quad (3)$$

where $[\mathbf{x}_t]_i \in \mathbb{R}$ is a sensor i reading from the time t , μ_i and σ_i are the mean and standard deviation of the all sensor i readings in this sequence.

3 Experiments

Our experiments use two sets of a gesture data to test the recognition rate of a HMM-based detector. The impact of the starting point selection and number of training sequences on the detection quality is provided.

3.1 Data Sets

The data for the experiments was captured with the DG5VHand motion capture glove, containing five finger bend sensors (resistance type), and three-axis accelerometer producing three acceleration and two orientation readings. The time between each readings ($t_{i+1} - t_i$) was approximately 30ms.

Two data sets were used in the experiments. The first one was a test set of gestures designed to examine the properties of gesture execution among different persons and with different movement speeds. The gestures are detailed in table 1. For discussion on gesture choice the reader is referred to [4]. There are 22 gestures in the set, each performed by four persons five times (three times normal speed, one fast, one slow).

During the work, it was noticed that in order to achieve a satisfactory level of detection, a larger training set may be needed. Since the current version of the database contains only $n = 20$ repetitions of every gesture, an additional testing data was created. Gesture numer 16 (*Grasping*) was selected for the experiment and a new set of $n = 40$ repetitions performed by one person was prepared using the same motion capture equipment that was used to create an original database.

3.2 Methodology

Our main objective is to observe the performance (measured with the detection rate γ) as a function of number of ‘teaching’ sequences, for different parameters (number of states, number of symbols, used preprocessing scheme, used dataset, HMM topology).

In all cases we use a single HMM for every gesture. Given a dataset with k gestures, for each gesture i we have a set of sequences $\mathcal{S}^{(i)} = \{\mathbf{S}_j^{(i)}\}$. We perform an experiment with the steps outlined below:

1. Initialize $n_r = 0$ (number of recognised sequences).
2. For every gesture, select previously unused sequence from $\mathcal{S}^{(i)}$ as test sequence, labelled $\mathbf{T}^{(i)}$. If every sequence has been used as a test, terminate. Set number of teaching sequences n_t to 1.
3. Identify HMMs for each gesture, using n_t of sequences remaining in $\mathcal{S}^{(i)} \setminus \mathbf{T}^{(i)}$.
4. Test recognition of each of the $\mathbf{T}^{(i)}$ with given HMMs; add the number of properly recognised sequences to n_r .
5. Increase n_t . If n_t is greater than number of sequences available to train, goto 2, else goto 3.

The γ is computed by dividing n_r by the number of tests done.

Table 1 The gesture list used in experiments (a part of this set was also used in [2])

Name	Class ^a	Motion ^b	Comments
1 <i>A-OK</i>	symbolic	F	common ‘okay’ gesture
2 <i>Walking</i>	iconic	TF	fingers depict a walking person
3 <i>Cutting</i>	iconic	F	fingers portrait cutting a sheet of paper
4 <i>Showe away</i>	iconic	T	hand shoves away imaginary object
5 <i>Point at self</i>	deictic	RF	finger points at the user
6 <i>Thumbs up</i>	symbolic	RF	classic ‘thumbs up’ gesture
7 <i>Crazy</i>	symbolic	TRF	symbolizes ‘a crazy person’
8 <i>Knocking</i>	iconic	RF	finger in knocking motion
9 <i>Cutthroat</i>	symbolic	TR	common taunting gesture
10 <i>Money</i>	symbolic	F	popular ‘money’ sign
11 <i>Thumbs down</i>	symbolic	RF	classic ‘thumbs down’ gesture
12 <i>Doubting</i>	symbolic	F	popular Polish(?) flippant ‘I doubt’
13 <i>Continue</i>	iconic ^c	R	circular hand motion ‘continue’, ‘go on’
14 <i>Speaking</i>	iconic	F	hand portraits a speaking mouth
15 <i>Hello</i>	symbolic ^c	R	greeting gesture, waving hand motion
16 <i>Grasping</i>	manipulative	TF	grasping an object
17 <i>Scaling</i>	manipulative	F	finger movement depicts size change
18 <i>Rotating</i>	manipulative	R	hand rotation depicts object rotation
19 <i>Come here</i>	symbolic ^c	F	fingers waving; ‘come here’
20 <i>Telephone</i>	symbolic	TRF	popular Polish(?) ‘phone’ depiction
21 <i>Go away</i>	symbolic ^c	F	fingers waving; ‘go away’
22 <i>Relocate</i>	deictic	TF	‘put that there’

^a We use the terms ‘symbolic’, ‘deictic’, and ‘iconic’ based on McNeill & Levy [6] classification, supplemented with a category of ‘manipulative’ gestures (following [8]).

^b Significant motion components: T-hand translation, R-hand rotation, F-individual finger movement.

^c This gesture is usually accompanied with a specific object (deictic) reference.

3.3 Results

Based on early experiments, best values were determined for the number of quantisation symbols $m \approx 60$ and number of HMM states $n \approx 5$.

Number of VQ symbols

Number of symbols m is one of the basic parameters of the VQ, its impact on γ is presented in figure 2 (left). Since this parameter governs VQ precision, high m value

results in very different sequences for multiple repetitions of the same gesture, hence a value of $m \approx 60$ was chosen in the experiments.

Number of HMM states

Figure 2 (right) presents the impact of the number of HMM states on γ . Its change does not significantly affect results, but the best performance was observed for $n \approx 5$, and it was used in experiments.

Topology and datasets

One of the basic steps to create an optimal detector is to choose a correct topology. Figure 3 presents the impact of chosen topologies (with and without an absorbing state) on detection results. ‘Ergodic’ and ‘Bakis’ topologies seem to yield best results for gesture data, but the detection rate of $\gamma \approx 0.62$ is not satisfactory (although the detector quality is clearly highly dependent on the number of training sequences).

The impact of a number of training sequences on γ for the selected ‘Grasping’ gesture (second dataset) is presented in Figure 4 (right). The satisfactory detector quality $\gamma \approx 0.96$ is achieved for $n \approx 20$ teaching sequences with the best result achieved using an ‘Ergodic’ HMM topology. The performance for the second dataset may result from the fact that it was recorded by one person without an intended change in motion speed.

Preprocessing

To test the impact of data preprocessing on the detector quality, the data was re-sampled and normalised. Results for raw and preprocessed data set are presented in Figure 4 (left). Basic preprocessing doesn’t significantly improve the detector quality although the resampled data may be used without the significant loss of detection quality.

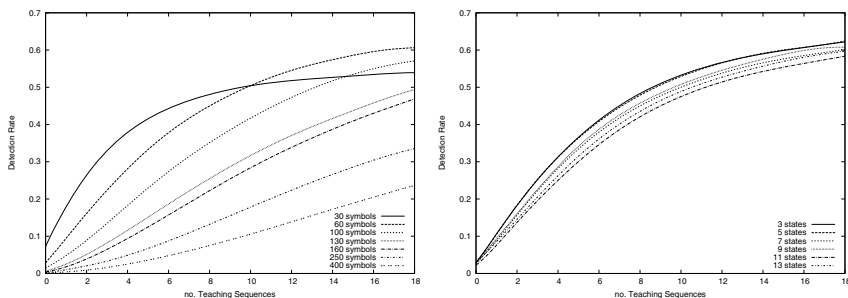


Fig. 2 The impact of classifier parameters on the γ . No. symbols for $n = 5$ states (left) and number of HMM states for $m = 60$ symbols (right)

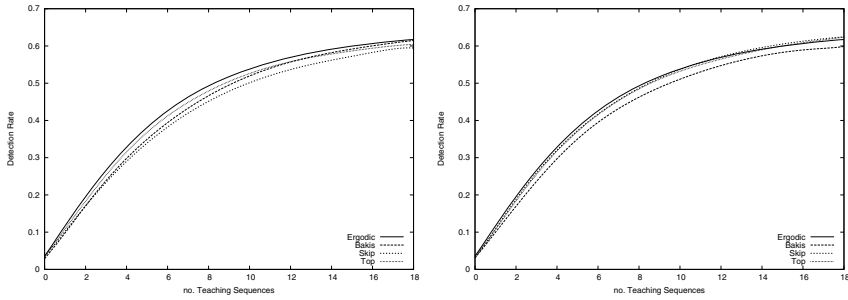


Fig. 3 The impact of the HMM topology on the γ , models without absorbing state (left) and with absorbing state (right). Number of states $h = 5$, number of symbols $m = 60$

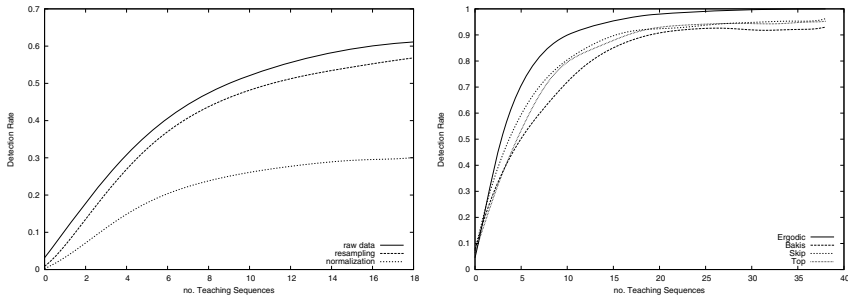


Fig. 4 The impact of data preprocessing on the γ for the gesture database and Bakis-topology HMMs (left) and the impact of the HMM topology on the classification results for the 'Grasping' gesture (right). Number of states $h = 5$, number of symbols $m = 60$

3.4 Discussion

On the basis of performed experiments it can be concluded that the simple HMM-based classifier using one model per gesture, trained with multiple sequences may not be sufficient for an effective gesture recognition without an additional data analysis. While the solution achieved a detection rate of $\gamma \approx 0.96$ using ($n \approx 20$) training sequences for a gesture data set created by one person, the performance for more diversified database with gestures performed by multiple persons with different speed was not satisfactory ($\gamma \approx 0.62$).

Since raw data of multiple gesture repetitions may be very different (eg, accelerometer sensor may not register a sufficiently slow gesture), an initial extraction of specific gesture features to be used for quantisation and/or a data dependent initialisation, may improve the detection rate for diversified databases.

While a choice of a correct number of VQ symbols is obviously essential, a relative tolerance to the selection of starting classifier hmm-related parameters (eg, adopted topology, the number of states) may be observed. This may mean that the classifier can maintain the similar detection quality for data from different motion capture devices.

4 Conclusion

Raw data HMM-based classifier may be a sufficient solution for gesture recognition in simple databases but the diversity of a gesture data requires an additional analysis to achieve a satisfactory performance. The solution seems relatively tolerant for starting parameter selection.

Acknowledgements. This work has been partially supported by a Polish Ministry of Science and Higher Education project NN516405137 ‘User interface based on natural gestures for exploration of virtual 3D spaces’. Authors would like to thank reviewers for their constructive comments.

References

- [1] Axenbeck, T., Bennewitz, M., Behnke, S., Burgard, W.: Recognizing complex, parameterized gestures from monocular image sequences. In: 8th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2008, pp. 687–692 (2008)
- [2] Cholewa, M., Głomb, P.: Gesture data modeling and classification based on critical points approximation. Accepted for 7th International Conference on Computer Recognition Systems CORES 2011 (2011)
- [3] Elmezain, M., Al Hamadi, A., Appenrodt, J., Michaelis, B.: A hidden markov model-based continuous gesture recognition system for hand motion trajectory. In: ICPR, pp. 1–4 (2008)
- [4] Głomb, P., Romaszewski, M., Opozda, S., Sochan, A.: Choosing and modeling gesture database for natural user interface. Accepted for 7th International Conference on Computer Recognition Systems CORES 2011 (2011)
- [5] Linde, Y., Buzo, A., Gray, R.: An algorithm for vector quantizer design. *IEEE Transactions on Communications* 28(1), 84–95 (1980)
- [6] McNeill, D.: *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press, Chicago (1992)
- [7] Elisabetta Farella Piero Zappi, Bojan Milosevic and Luca Benini. Hidden markov model based gesture recognition on low-cost, low-power tangible user interfaces. *Entertainment Computing*, 1(2):75–78, 2009.
- [8] Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., McCullough, K.E., Ansari, R.: Multimodal human discourse: gesture and speech. *ACM Trans. Comput.-Hum. Interact.* 9, 171–193 (2002)
- [9] Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 257–286 (1989)
- [10] van Oosten, J.-P.: Can markov properties be learned by hidden markov modelling algorithms? Master’s thesis, University of Groningen, The Netherlands (2010)
- [11] Yang, H.-D., Park, A.-Y., Lee, S.-W.: Human-robot interaction by whole body gesture spotting and recognition. In: International Conference on Pattern Recognition, vol. 4, pp. 774–777 (2006)
- [12] Yang, J., Xu, Y.: Hidden markov model for gesture recognition. Technical Report CMU-RI-TR-94-10, Robotics Institute, Pittsburgh, PA (May 1994)

A Generic Approach to the Texture Detection Problem in Digital Images

Christian Feinen, Marcin Grzegorzek, Detlev Droege, and Dietrich Paulus

Abstract. In this paper, we describe our solution to the marker detection problem in digital images. In order to keep our investigations as general as possible, our approach has not been developed in an application-driven way. However, we have evaluated the system for the bar code detection problem. To represent the markers, our system uses general feature extraction methods like *Hu Moments* and the *Fourier-Mellin* transform which are both invariant to rotation, scaling and translation. For marker classification, *Bayes Classifier* and *Support Vector Machine* have been applied. A comprehensive set of experiments performed for our algorithm proved its high robustness for a challenging set of images.

1 Introduction

Numerous image processing applications rely on the detection of characteristic visual structures in unknown regions of input images. This may be the detection of 1D or 2D bar codes on commercial goods, the detection of license plates, traffic signs or landmarks in image series as well as the navigation of autonomous robots. Usually, regions of interest may appear anywhere in an image and should be identified as accurate as possible to be worthwhile for further thorough analysis. Most of such regions can be considered to represent a texture with unique description parameters. Searching for such characteristic textures based on invariant features is a promising approach for a universal component identifying such regions of interest.

In this paper, we propose a robust solution to this problem in form of a generic application-independent approach by combining well established invariant

Christian Feinen · Marcin Grzegorzek
Research Group for Pattern Recognition,
University of Siegen, Germany

e-mail: [{{christian.feinen,marcin.grzegorzek}@uni-siegen.de](mailto:{christian.feinen,marcin.grzegorzek}@uni-siegen.de)

Detlev Droege · Dietrich Paulus
Research Group for Active Vision,
University of Koblenz-Landau, Germany

e-mail: [{{droege,paulus}@uni-koblenz.de](mailto:{droege,paulus}@uni-koblenz.de)

measures like Hu moments [6] and the Fourier-Mellin transform [13] with supervised classification approaches, namely the Bayes Classifier and the Support Vector Machine. In this way, an efficiently adaptable and reconfigurable component for ROI detection can be provided.

The paper is structured as follows. In Section 2, an overview of existing related approaches in this research field is given. Section 3 presents our detection method with its feature extraction and classification parts. In Section 4, a comprehensive set of experiments, which have been performed for the quantitative evaluation of the system, is presented. Section 5 closes the paper by providing some interesting conclusions, final remarks, and research plans for the future.

2 Related Work

The image processing and pattern recognition community has been working on the problem of detecting characteristic textures in digital images for years. However, most of the systems have been developed for a certain application domain. In general, they are not easily generalisable, i.e., they can not be applied for different applications without significant changes. This is the main feature differentiating our approach from the state-of-the-art.

For example, there are systems developed specifically for the detection of planar markers (e.g., 2D bar codes). Many of them work in a very similar way. First, they apply a simple thresholding approach followed by finding connected groups of pixels [4]. Other methods like those described in [2, 4] make use of well known edge detectors in order to find regions of interest. For this, statistical or gradient approaches as well as the Hough transform (see [1, 12]) are often applied. Other solutions focus on the specific code pattern and are usually faster than the former techniques [11]. Furthermore, region growing algorithms and texture analysis procedures [5] are also under focus in this context. Many methods employ numerous preprocessing steps like skew angle estimation based on the Fourier transform, Hough transform, directional projection or least square method (LSM) [8]. Template matching algorithms can also be found in the literature for finding 2D bar codes in digital images. In [10] such a technique using a Normalised Cross Correlation (NCC) is described.

3 Detection

As mentioned above, in this paper we describe a method for planar marker detection. Our approach uses two different types of features which are both invariant to rotation, scaling and translation. We will abbreviate this expression with the term “RST” in the following text. By using this special feature type we could reduce the false-positive rate of marker detection and avoid an extra preprocessing step as described in Section 2. In this paper, we only measure the detection rate, not the recognition success between different markers. The latter challenge is not solved with our proposed methods, but is performed by subsequent processing.

The employed features are the *Hu Moments* and *Fourier-Mellin* values. The former belong to the category of “texture features” and describe a pattern based on its intensity values. The latter operate on a Fourier and a log-polar transform to produce an invariant Fourier spectrum. The *Hu Moments* have been introduced by Ming-Kuei Hu in 1962. They are based on non-orthogonal centralised moments (see equation (I)),

$$\mu_{pq} = \iint_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q I(x, y) dx dy \tag{1}$$

$$\text{with } \bar{x} = \frac{m_{10}}{m_{00}} \quad \text{and} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

which are translation invariant and can be easily made scaling invariant by performing a normalising step as shown in equation (2). In order to achieve the desired rotation invariance M.-K. Hu employed the theory of algebraic invariants to reformulate the centralised moments. As a result of these new expressions Hu was able to derive his seven RST invariant features of 2D objects [6]. Table 1 shows the mathematical formulation of groups of nonlinear centralised moment terms.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \gamma = \frac{p + q + 2}{2} \tag{2}$$

Table 1 Mathematical formulation of Hu’s seven invariant moments [6]

ϕ_1	$\eta_{20} + \eta_{02}$
ϕ_2	$(\eta_{20} - \eta_{02})^2 + 4(\eta_{11})^2$
ϕ_3	$(\eta_{30} - 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2$
ϕ_4	$(\eta_{30} + \eta_{12})^2 + (\eta_{03} + 2\eta_{21})^2$
ϕ_5	$(\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] + (3\eta_{21} - \eta_{03})(\eta_{03} + \eta_{21})[3(\eta_{30} + \eta_{12})^2 - (\eta_{13} + \eta_{21})^2]$
ϕ_6	$(\eta_{20} + \eta_{02})(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21})$
ϕ_7	$(3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] + (3\eta_{12} - \eta_{30})(\eta_{03} + \eta_{21})[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2]$

The next feature type that shall be considered in this paper is called *Fourier-Mellin*. Feature values of this type can be calculated by performing a Fourier and a Mellin transform. Since the Fourier transform is a well-known operation we don’t go too much into detail here. The most important property of the Fourier transform in this context is its natural translation invariance to all 2D object displacements. On the other hand, the Mellin transform leads to a 2D scaling invariance. Since the Mellin transform is not as popular as the Fourier transform we provide a short description. As already mentioned, the resulting values of the Mellin transform are scaling invariant. More precisely, it is an integral transform of a real- or complex-valued function $f(x)$ and it is defined as shown in (3) [13].

$$MT\{f(x)\} = M(k) = \int_0^\infty f(x)x^{-ik-1} dx \tag{3}$$

However, the practical use of this operation is limited due to the fact that the native use of the sampling theorem is not possible in contrast to the Fourier transform. Furthermore, there are no algorithms available performing a fast Mellin transform. Thus the general process for computing such values prevents the direct use of the Mellin transform. Fortunately, it is possible to map the definition of Mellin to a Fourier operation by logarithmic scaling the Mellin input data as shown in [13]. Accordingly, the Fourier-Mellin process pipeline changes as depicted in Figure 1. Since we only achieve translation and scaling invariance by performing a Fourier

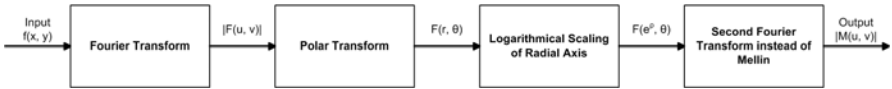


Fig. 1 FM Pipeline for producing RST invariant features (taken out from [13])

and a Mellin transform we need a further step, the polar transform, to obtain RST invariant features. The main advantage associated with this operation is the decoupling of rotation and scaling simply by defining the spectral magnitudes in polar coordinates. Within the polar representation, rotation is reduced to a translation and appears as a phase shift within the spectrum now [7].

By performing the process pipeline shown in Figure 1 we receive a 2D spectrum of the size of the input image or of an optimal size correlated to the Fourier transform. It is not important which one is chosen, because the feature space is too large in both cases. Thus, a PCA for reducing the feature space has to be performed. In addition to this, high pass filtering is necessary for sorting out all dominant low frequencies.

Finally, we want give some information about the classifiers used. In this paper we compare two classifiers combined with the features introduced above. The classifiers used are the *Bayes Classifier* and the *Support Vector Machine*. The former is a probabilistic based classifier which is useful for one- and multi-class-problems. The latter is a non-probabilistic linear classifier. Traditionally, the Support Vector Machine was designed for classifying two-class problems. Today, there are implementations that are equally capable for solving one- or multi-class problems. Here we use a Support Vector Machine specified for two-class problems, with the aim to evaluate this case combined with a one-class problem. To obtain a detail overview about the mentioned classifiers, please refer to the book “Pattern Recognition” [14] by Theodoridis and Koutroumbas.

These classifiers can be parametrised in different ways. For example, to estimate the probability density function of a pattern that is used by the Bayes classifier you have to choose how to sample the function. This can be done by a parametrised approximation like a Gaussian distribution function or in a non-parametrised way by using a histogram. We have used the latter method. In addition to this, we have

to set a threshold because we want to solve a one-class problem and therefore we cannot compare *a posteriori probabilities* of different classes with each other. The implementation of the Support Vector Machine supports equally different parameters which can be used to optimise the classification process. These parameters include *kernel functions*, a *scaling factor*, *functions to find the separating hyper-plane* and many more. For our purposes we use the default parameter set provided by Matlab.

4 Evaluation

In this section, we describe our experimental setup to evaluate the features and classifiers introduced in section 3 for a bar code detection problem. In addition to this, we involve a third feature type which is not totally RST invariant, namely the *Haralick Texture Features*. They belong to the category of texture features similar to the Hu Moments. In its core version, this feature type consists of thirteen values which describe the corresponding texture. By using the implemented parameters you can compute the values for different directions. For obtaining further information about the Haralick Features, please refer to [3]. For our purposes we configure the Haralick Features with two different parameters. The first one determines the distance between pixels which are used to compute the co-occurrence matrix; we set this factor to one (for all directions). The second parameter determines the angle for what direction the values shall be computed; here the values are estimated for 0° , 45° , 90° and 135° .

We organised the evaluation process as a 10-fold cross-validation. This means, that the whole set of available data (here: 2D bar codes) is divided into 10 subsets. Thus, one evaluation sequence consists of ten iterations. For each iteration nine subsets are taken to train the classification system and one subset is used to test it. Altogether, this is done ten times with each subset successively. Overall, we have 3000 codes of size 342×342 pixels and, therefore, each subset consists of 300 codes. As test pattern we take an Aztec-Code (see figure 2(a) [9]). Furthermore, our purpose is only to detect the code in a picture and not to analyse it so the choice of the pattern is not that important as long as it is similar to a marker. In each iteration, all codes are placed randomly in an arbitrary picture that is used as search area. Thereby, 90 codes are randomly scaled and 90 codes are rotated by a random angle; the remaining 120 codes only differ in their position within the picture. The test pictures for searching are captured with an ordinary digital camera and have a size of 1600×1200 pixels. Their background is heterogeneous showing different patterns, objects and landscapes. In order to check quality requirements of the input data, all configurations leading to good results are performed once again, but this time a two-step Gaussian noise is added to the input data.

Once we have prepared our subsets and merged search areas with bar codes we train our classification system and test it. In relation to this we need a search strategy which consequently affects both, the training and the search process. In this paper,

we do not use a traditional sliding window since its handling is problematic. For example, it is difficult to consider things like overlapping patterns or size changes. Our method is designed as follows. First, we slice the pattern into configured sub regions and then we train our classifier with all of them. After that and second, we partition all test images (merged search areas) into the described sub regions and present them to the classifier. It takes the decision whether a sub region is part of a pattern or not (see figure 2(b)). This approach also provides the possibility

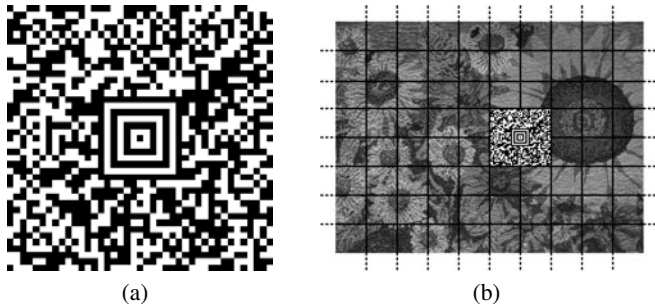


Fig. 2 Left: Aztec-Code Example; **Right:** Merging example of search area and code

to estimate the pattern size or to detect occluded patterns. The values of the Hu Moments are computed as shown in Table 1. Since the values are close to zero and quite different, it is useful to work with the logarithms of their absolute values as demonstrated in equation (4).

$$\phi'_n = \log(|\phi_n|) \quad \forall \phi_n \neq 0, \quad n = 1, \dots, 7 \quad (4)$$

In context of the Fourier-Mellin features the best results can be achieved with 6 to 15 eigenvectors. We also performed tests with 30 eigenvectors revealing that this number is too large to discriminate the searched pattern from the background. The evaluation environment uses *Matlab* on an UNIX system (Ubuntu 64bit) with an Intel Core 2 Duo processor and 4GB RAM. We employ predefined functions like `fft2`, `cov` or `eig` to compute all needed values. To interpret the evaluation results it is necessary to know what kind of tests we are running. First of all, we test the detection success rate, which is one of the most important factors for such a system. Secondly, we investigate the false-positive rate influenced by the amount of regions which are wrongly classified as searched patterns. Since we work on sub regions we need a neighbourhood constraint that tells us, if there are enough connected sub regions to build a pattern. This amount correlates with the size of the currently merged code that varies permanently. If a bar code has its normal size of 342×342 pixels which equals to 3×3 sub regions, the classifier has to find five of them to mark this region as a found pattern. In contrast to this, if there are only four connected sub regions which are wrongly indicated as sub patterns the false-positive rate increases

by one. The detection success rate (DR) is a number between 0 and 1, whereas 1 means that all bar codes has been found. However, the false-positive rate (FPR) is the sum of all misclassified regions. Both values, the detection success rate and the false-positive rate are mean-values of the 10-fold cross-validation. We also provide the average calculation time (ACT) which is needed for one set of 140 sub regions.

As already stated, we compute a threshold value. This is done during the training phase of the Bayes Classifier. To evaluate this value we multiply it by a “threshold factor” to increase or to decrease it.

4.1 Results of Bayes Classifier

4.1.1 Configuration: 30 Bins and 0.8 Threshold Factor

Feature Type	DR	FPR	ACT
Hu Moments	0.9790	25.9580	0.3016
FM Features (6 EigenVecs)	0.8197	0.1263	2.2479
FM Features (15 EigenVecs)	0.9517	1.5670	2.3091
FM Features (15 EigenVecs, noisy)	0.9990	67.9060	2.0535
Haralick Texture Features	0.0177	0	19.7260

4.1.2 Configuration: 30 Bins and 1.0 Threshold Factor

Feature Type	DR	FPR	ACT
Hu Moments	0.7427	5.7537	0.3009
Hu Moments (noisy)	0.7417	5.8873	0.2784
FM Features (6 EigenVecs)	0.3490	0.0030	2.2620
FM Features (15 EigenVecs)	0.6033	0.1233	2.3215
Haralick Texture Features	0.0070	0	19.9680

4.1.3 Configuration: 60 Bins and 0.8 Threshold Factor

Feature Type	DR	FPR	ACT
Hu Moments	0.9803	24.1790	0.3041
FM Features (6 EigenVecs)	0.8057	0.0917	2.2363
FM Features (15 EigenVecs)	0.9130	1.2700	2.3106
Haralick Texture Features	0.0203	0	18.9708

4.1.4 Configuration: 60 Bins and 1.0 Threshold Factor

Feature Type	DR	FPR	ACT
Hu Moments	0.6920	4.9407	0.3030
FM Features (6 EigenVecs)	0.3080	0.0010	2.2673
FM Features (15 EigenVecs)	0.1527	0.0153	2.3558
Haralick Texture Features	0.0093	0	19.1208

4.2 Results of Support Vector Machine

Feature Type	DR	FPR	ACT
Hu Moments	0.2887	12.6367	0.3925
FM Features (6 EigenVecs)	0.8677	0.1517	2.6906
FM Features (15 EigenVecs)	0.8757	0.2667	2.4151
FM Features (15 EigenVecs, noisy)	0.8333	46.8997	2.1307
Haralick Texture Features	0.9777	0.0027	21.3422
Haralick Texture Features (noisy)	0.8153	8.1083	52.9953

After performing all of our tests we are able to derive the following conclusions. Firstly, the Fourier-Mellin transform leads to a set of features which are more effective than the values of the Hu Moments. These are not able to discriminate the searched patterns sufficiently against the heterogeneous background of our test images. Secondly, the Haralick Features provide a high and accurate detection rate and their FPR values are really impressive. However, the calculation time as well as the detection time needed by the Haralick Features are a substantial drawback. A similar problem is observable in context of the Fourier-Mellin transform. Due to the dimension problem and the PCA the training phase is very time consuming. In general, this step can be performed offline and a detection time around 2.5 seconds is acceptable for a number of applications. Further observations which are derivable from our evaluation relate to the classification systems. Both, the Bayes Classifier and the Support Vector Machine, produce high detection rates. Only the combination of a SVM and Hu Moments leads to a disappointing result. In context of the Bayes Classifier an appropriate threshold must be chosen to achieve a high detection rate as it is not the case for the Haralick Features in our evaluation. The value of this threshold is not obvious and an application-driven threshold can be chosen experimentally. Finally, it is interesting to know that the Hu Moments are quite robust against noisy input data.

5 Conclusion

As already mentioned in Section 1, detection of characteristic textures is an essential step in many image processing systems. It is important to improve solutions to the detection problem with the aim to limit the search space of subsequent processing procedures and to avoid unnecessary work. In this paper we presented a generic approach to the problem of texture detection using invariant features. In contrast to the most existing approaches from this research field, our system has not been developed for a certain application domain (Section 2). Although we evaluated the system for the 2D bar code detection problem, generally it can be applied for other kinds of textures as well. Thus, there is a lot of potential to adapt our methods in an application-driven way. In order to extract features for the textures to be detected, our system computes *Hu Moments* and values of a *Fourier-Mellin* transform which are both invariant to rotation, scaling and translation. This is an enormous advantage in comparison to other feature types like the Haralick Features, since time consuming preprocessing during the detection phase has been omitted in this way; e.g. it is not necessary to rotate the sliding window above the same position for different angles and to compare all of the resulting values with the one generated during the training phase. For the classification, supervised pattern recognition techniques like the Bayes Classifier and the Support Vector Machine are used (Section 3). In Addition to this, we have performed a comprehensive set of experiments (Section 4) and achieved promising results for the 2D bar code detection problem. From the perspective of the detection rate there is no difference between the investigated classifiers. However the Bayes Classifier is more suitable to work with less discriminatory features like the *Hu Moments*.

In the future, further feature extraction methods and classification schemes will be integrated into our framework. There are numerous invariant feature types which are able to solve the addressed problem. Moreover, we will investigate unsupervised methods in this context. Additionally, we are going to compare the results of the two-class SVM with the detection rate of an one-class SVM. Finally, the evaluation results will be proved in an application-driven domain.

Acknowledgements. This work was funded by the German Research Foundation (DFG) as part of the Research Training Group GRK 1564 “Imaging New Modalities”.

References

- [1] Duda, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM* 15, 11–15 (1972)
- [2] Fiala, M.: Artag, a fiducial marker system using digital techniques. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. 2, pp. 590–596 (2005)
- [3] Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3(6), 610–621 (1973)

- [4] Hirzer, M.: Marker detection for augmented reality applications. Tech. rep., Institution for Computer Graphics and Vision, Graz University of Technology, Austria (2008)
- [5] Hu, H., Xu, W., Huang, Q.: A 2d barcode extraction method based on texture direction analysis. In: Fifth International Conference on Image and Graphics, ICIG 2009, pp. 759–762 (2009)
- [6] Hu, M.K.: Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* 8(2), 179–187 (1962)
- [7] Jin, A.T.B., Ling, D.N.C.: Integrated wavelet and fourier-mellin invariant feature in fingerprint verification system. In: Proceedings of the ACM SIGMM Workshop on Biometrics Methods and Applications, WBMA 2003, pp. 82–88. ACM, New York (2003)
- [8] Liang, Y.H., Wang, Z.Y.: A skew detection method for 2d bar code images based on the least square method. In: 2006 International Conference on Machine Learning and Cybernetics, pp. 3974–3977 (2006)
- [9] Longacre Jr., A., Hussey, R.: Two dimensional data encoding structure and symbology for use with optical readers (1997), U.S. Patent No. 5,591,956 (“Aztec-Code”)
- [10] Nakthanom, S., Choomchuay, S.: A 2d barcode inspection using template matching. In: Proceedings of the 3rd International Conference on Data Storage Technology (DST-CON 2010) (2010)
- [11] Parvu, O., Balan, A.G.: A method for fast detection and decoding of specific 2d barcodes. In: Proceedings of the 17th Telecommunications forum TELFOR 2009, pp. 1137–1140 (2009)
- [12] Rau, J.Y., Chen, L.C.: Fast straight lines detection using hough transform with principal axis analysis. *Transform* 8(1), 1–17 (2003)
- [13] Schalkoff, R.J.: *Digital Image Processing and Computer Vision*. Wiley, Chichester (1989)
- [14] Theodoridis, S., Koutroumbas, K.: *Pattern Recognition*, 4th edn. Academic Press, London (2009)

Waterball - Iterative Watershed Algorithm with Reduced Oversegmentation

Michał Swiercz and Marcin Iwanowski

Abstract. In this paper we present a new approach to watershed algorithm for segmentation of digital grey-scale images. The approach is derived from rainfall-type watershed methods, but utilises a different method of path tracing and iterative gradient image preparation to reduce oversegmentation and yield better results in object extraction. Sample results are discussed, with emphasis on their global correctness and practical applications.

1 Introduction

The watershed transformation is an established, powerful tool for morphological image segmentation. It has found many applications in different areas of science. The concept of watersheds comes from the field of topography, referring to the division of a continental landscape into water catchment basins, which collect rain water flowing down from areas of land associated with them. This concept can be utilised in image segmentation to simulate this water behaviour and generate partitions of the image corresponding to objects of interest. A grey-scale picture is considered a topographic relief, where the height of a point is equal to the grey level of the corresponding pixel.

The watershed segmentation method is not usually applied directly to the original intensity image, as this will not produce meaningful segmentation results for most images. We are interested in extracting objects out of the background that are most homogeneous in respect of their grey level. High values of pixel gradient usually correspond to object boundaries, so when considering the gradient image as a topographic relief, these boundaries should form catchment basins corresponding to objects of interest. Therefore, the gradient image is much more useful as the base

Michał Swiercz · Marcin Iwanowski
Institute of Control and Industrial Electronics,
Warsaw University of Technology
ul.Koszykowa 75, 00-662 Warsaw, Poland
e-mail: swierczm,iwanowski@ee.pw.edu.pl

input material for watershed segmentation processing. The computation of gradient magnitude is quite straightforward, though many variations exist e.g., Sobel, Prewitt, or morphological gradient operators.

During the recent years, two conceptually distinct techniques have been developed to calculate the watershed transform. The first method proposes the transformation by immersing the topographic surface in a lake, with holes pierced in all local minima of the surface. This technique was implemented in e.g. [4, 12]. The second method simulates the behaviour of rain water over the topographic surface associated with the image, so is in a sense the reverse of the immersion approach. This method has been described, amongst others, in [5, 8, 9, 10].

One of the main drawbacks of the classical watershed algorithm is a phenomenon known as oversegmentation. The watershed transform tends to properly extract the edges of the objects of interest, but break up their interior into many small segments, as each local minimum of the gradient image becomes a seed of a new region, even though these minima are irrelevant in the global scope. There exist several techniques of countering oversegmentation, most commonly based on blurring the input gradient image. This, however, creates the risk of loosing fine detail and disrupting object boundaries.

The „waterball” method presented in this paper can be perceived as a composite method for object extraction, combining several techniques and mechanisms to produce satisfactory macro-scale results and demonstrating an alternative approach to the oversegmentation problem in watersheds. The method was designed to counter oversegmentation during the actual calculation of watersheds, without the need to heavily blur the image beforehand, which can result in rapid, undesired loss of detail. Also, there is no post-processing phase, where region merging is performed - the partitioning is obtained solely by modifying the input image, calculating watersheds and evaluating results.

The „waterball” method uses two distinct mechanisms of reducing oversegmentation. First of these mechanisms, the "rolling ball", was based on a simulation of a larger body, such as a massive ball, rolling down the slope. Contrary to classic rainfall-type watershed methods, where a single-pixel drop of water is used, this ball has the ability to cross small ridges and ignore small, insignificant local minima. When such a minimum is encountered, the analysis of extended neighbourhood is performed, and if a certain type of barrier is encountered, it can be crossed in search of a more dominant minimum. The second mechanism, weakening with edge enhancement, makes it possible to eliminate low, insignificant interior boundaries, which contribute to oversegmentation. In each iteration of the algorithm, a new gradient image is constructed, by means of first weakening the one used in previous iteration, and then strengthening edges to prevent detail loss.

The paper is organized as follows: in section 2 we provide a detailed description of our basic algorithm for image segmentation. Some fundamental notations are defined and the concept of „rolling ball” is outlined. In section 3, we present an edge enhancement method and finally, an iterative image segmentation algorithm (the „waterball” algorithm) utilising the „rolling ball” mechanism and edge enhancement

to produce image partitionings with low oversegmentation. Results, as well as merits and drawbacks of our method are discussed in section 4, together with their impact on practical applications. Section 5 concludes the paper.

2 Base Segmentation Algorithm

2.1 Basic Definitions and Notations

Before describing in detail our approach, we introduce some basic definitions used throughout this paper. We define the following terms:

1. **Direction** is any vector spanning between two adjacent nodes of underlying connectivity grid, with an emphasized origin node. Therefore, with 4-connectivity, there exist four directions associated with a given pixel (which can be intuitively understood as *left*, *right*, *up* and *down*), with 8-connectivity, there exist additional four directions (*left-up*, *left-down*, *right-up* and *right-down*).
2. Intensity function of pixel p will be denoted as $f(p)$. The value of the intensity function may be intuitively associated with the height of the pixel on the topographic relief.
3. Pixel q is considered a **neighbour** of pixel p if and only if pixel q is adjacent to pixel p on the underlying connectivity grid. If the height of q ($f(q)$) is the lowest of all the neighbours of p , and also lower than the height of p ($f(p)$), q is the **lowest neighbour** of p . If there are multiple pixels meeting this criterion, they are all considered lowest neighbours of p . Also, in regard of the previous point, q is a neighbour of p in a certain **direction**, if p is the origin point for this direction.
4. **Path** is a series of pixels, in which every two consecutive pixels in the series are neighbours in the image, and every pixel in the series is unique. As **length** of path we will understand the number of pixels in the path.
5. Pixel p_B is a **jump neighbour** of pixel p_A in direction D if all of the below conditions are met:
 - a. Pixel p_B is a neighbour of p_A in direction D .
 - b. Pixel p_B has a greater height than pixel p_A ($f(p_B) > f(p_A)$).
 - c. Pixel p_B has a neighbour p_C in direction D .
 - d. Height of p_C is lower than height of p_A ($f(p_A) > f(p_C)$). Pixel p_C will be called **jump target** of pixel p_A .
6. **Jump function** of pixel p_A in direction D is defined as follows:

$$J(p_A, D) = af(p_B) + bf(p_C), \quad (1)$$

where p_B is a jump neighbour of p_A in direction D and p_C is a jump target of p_A in direction D . Values a and b are constants chosen during the tuning phase

of the algorithm. Jump function for a pixel exists only if both jump neighbour and jump target exist for this pixel.

7. Pixel p_B is a **lowest jump neighbour** of pixel p_A if it is a jump neighbour of p_A , and the jump function of p_A associated with p_B (being A 's jump neighbour) is the lowest of all the jump functions of p_A .
8. **Plateau** of height h is a set of pixels such that all the pixels are of the same height h and every two pixels can be connected by a path with all the pixels in the path also having a height of h . Plateau pixel is considered an outlet, if it has at least one lower neighbour or at least one jump neighbour.
9. **Minor minimum** is a pixel that does not have a neighbour with a lower height but has at least one jump neighbour, or a plateau consisting of pixels without lower neighbours but having at least one jump neighbour.
10. **Major minimum** is a pixel that does not have a neighbour with a lower height or a jump neighbour, or a plateau consisting of pixels without lower neighbours or jump neighbours.
11. **Roll path** is the shortest possible path from pixel p_A towards pixel p_B , where each successive pixel is either the lowest neighbour of the previous pixel in the path, or a lowest jump neighbour of the previous pixel if the previous pixel has no lower neighbours. Additionally, the starting pixel p_A is not lower than pixel p_B . Every roll path ends in a major minimum but can cross several minor minima.

An additional explanation is required for the case of roll paths on plateaus. This is a special scenario, as inner pixels of a plateau do not have a lowest neighbour and cannot have jump neighbours. In this case, we assume geodesic distance [7] as distance criterion on the plateau and on this basis we calculate the shortest path between the plateau entry pixel and the nearest outlet.

As we mentioned before, there can exist several equally valid roll paths for a pixel, leading to different major minima. Therefore, such a pixel can belong to more than one catchment basin and a criterion must be chosen to determine which basin to use. In this work, we use the „first-found” criterion, by which to continue the roll path we pick the first viable lowest neighbour or, if not possible, the first lowest jump neighbour found during a pixel's neighbourhood scan. This is a simple and fast method that, even though is biased in certain specific scenarios, generates satisfying overall results and has virtually no meaningful impact in the macro-scale.

The jump function is one of the central concepts behind the waterball algorithm. Jump pixels can be considered a cross-section of the barrier that blocks the movement of the rolling ball. By analysis of the shape of that cross-section, it can be determined whether the barrier is likely to be a small obstruction due to noise or similar phenomena, or if it is a proper division area between meaningful objects in the image. Then, the significance of this barrier can be assigned a numeric value, which in turn can be associated with pixel height and in this respect compared with other pixel heights when determining further roll path.

2.2 Path Tracing by “Rolling Ball”

Our "rolling ball" algorithm implements the rainfall approach to find a downstream flow along the steepest roll path from each pixel, and distribute region labels marking the association of each pixel to a certain regional minimum. A mechanism called directional code is used to code roll paths, where each visited pixel receives a marking in the dedicated 'direction array' - a specially created integer array equal in size to the image array. The value of this marking represents the position of pixel in the image that should become the next step in the roll path. A path can be therefore viewed as a series of pointers to pixels stored in the direction array.

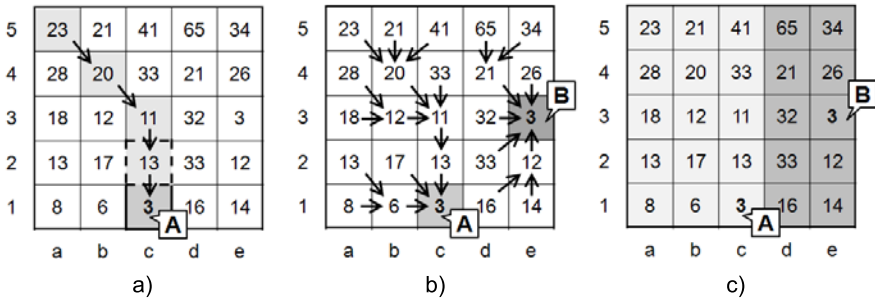


Fig. 1 Rolling ball segmentation: single roll path from pixel a5 (a), fully scanned image with all direction markings (b), and the final partition of the image (c)

First, the image is surrounded by a 2-pixel wide frame of *guardian pixels* which are of set, much higher altitude than any of the actual pixels in the image, and therefore provide an uncrossable boundary for roll path tracing. This is done to improve performance and eliminate unnecessary checks of boundary conditions in each neighbourhood scan. The path-tracing process is performed for each pixel in the image and consists of the following steps:

1. Neighbourhood scan is performed to find the lowest neighbour of the starting pixel. The lowest neighbour of the starting pixel becomes the next step in the roll path and is marked with directional code. Each new pixel in the path is in turn scanned for its lowest neighbour to add to the path, and the process continues until a minimum is encountered. This case is demonstrated on the example of pixels a5 or b4 in Fig. 1(a).
2. If a minimum is encountered, it is scanned to determine if it is a minor or a major minimum. If a jump pixel (neighbour) is found, this minimum is a minor minimum and this jump neighbour is treated as lowest neighbour and becomes the next step in the roll path. Path tracing continues at that jump pixel. If no jump pixels are found, this candidate minimum is considered a major minimum and every pixel of this minimum is labeled with a new label (unless it was already labeled during previous processing). In Fig. 1(a), pixel c3 is a minor minimum,

for which jump neighbours c_2 and e_3 exist. Jump neighbour c_2 is preferred (it is the lowest jump neighbour for c_3) and the roll path can continue from c_3 through c_2 to pixel c_1 , which is a major minimum. Pixel c_1 , as a newly encountered major minimum, receives a new label, „A”.

3. If a plateau is encountered, it is scanned to find all the pixels belonging to it, as well as to locate the plateau's outlets. Every outlet is analysed, and its next step is found (note that a lower neighbour always takes precedence before a jump neighbour when considering a next step). The plateau itself is then labeled using the region growing *prairie fire* method, where the outlets act as seeds for propagation and distribute their next step to their neighbours on the plateau. The next step markers spread with each iteration of the region growing to gradually fill the whole plateau. After spreading the next step markers over the plateau, the processing continues to the next pixel in the image.
4. After every pixel is given a direction marker (Fig. 11(b)), the whole image is scanned again, and for every pixel, its roll path is traced until an associated major minimum is found. Since all the major minima are labeled at this point, the appropriate label can be propagated along the processed roll path. At the end of this step, the whole image is labeled (Fig. 11(c)).

It is worth pointing out that there are many cases, where roll paths originating from different pixels merge at some point along the way. For example, in Fig. 11(a), paths originating from pixels a_4 and a_3 meet and merge at pixel b_3 . To avoid unnecessary computations, whenever an already labeled pixel is encountered during roll path tracing, the path is not traced further (as it was obviously already scanned and labeled) and we proceed with propagation of the encountered label, just as if a minimum was reached. Therefore, only one of the paths from pixels a_4 or a_3 will be fully retraced, the other will be retraced only from the origin point to the merge point.

3 Segmentation Algorithm with Iterative Edge Enhancement

3.1 Weakening with Edge Enhancement

The gradient image calculated from the original intensity image (for example using the classic Sobel method) usually captures object boundaries sufficiently well, representing them as high-intensity edges. However, often the interior of the objects is not uniform and some small internal boundaries also manifest in the gradient image. These boundaries are not globally significant and are a major cause of undesired oversegmentation. To reduce their impact on segmentation results, an immersion technique was designed, consisting of the following steps:

1. The gradient image is weakened by subtracting from it a constant, relatively small value. In this work, weakening value of 2 was used for gradient images with intensity range between 0 and 255. If there are pixels that have a negative intensity value after this subtraction, their value is rounded to zero.

2. Edge intensity image is calculated from the weakened gradient image obtained in step 1. In this work, Roberts method was used.
3. Edge intensity image, obtained in step 2, is multiplied by a scaling factor. In this work, a scaling factor of 1.1 was used.
4. Such prepared edge intensity image is added to the weakened gradient image from step 1 to produce a final gradient image. This final gradient image is the input material for segmentation.

The weakening technique allows to remove smallest internal boundaries, which are insignificant compared to the strong, external object edges, but would heavily influence the results of watershed transform. The edge enhancement helps mitigate the effect of weakening on main boundaries in the image. As will be demonstrated later, the described mechanism, together with rolling ball path tracing works very efficiently in the iterative segmentation method.

3.2 *Iterative Waterball Segmentation*

The iterative waterball segmentation utilises two techniques described in the previous sections: „rolling ball” segmentation and edge enhancement. The algorithm can be summarised as follows:

1. Initial gradient image is prepared from the intensity image
2. The waterball segmentation is performed on the gradient image to extract segment labels. Values $a = 2$ and $b = 1$ were used as jump function parameters.
3. Stop condition test is performed. If the stop condition is satisfied, processing ends. If not, execution proceeds to step 4.
4. Weakening with edge enhancement is performed on the gradient image, forming new input for the next iteration.
5. Execution returns to step 2 and the next iteration begins.

Due to the iterative nature of this method, a stop condition needs to be formulated. Optimally, the algorithm should yield results that are consistent with an intuitive partitioning of the image into macro-scale objects of interest. Therefore, in most cases, it is very hard to define the stop condition without human input. There exist several guided watershed algorithms, where human input is taken into account when performing the watershed segmentation, such as marker-based methods. However, it is sometimes difficult to pinpoint proper locations for markers. In this work, the concept of **dominant segment set** is introduced to function as a simple, intuitive stop criterion. Dominant segment set is a set of image segments, whose total pixel count equals or exceeds a predefined threshold (for example, 90% of a total pixel count of the image). The number of segments belonging to the dominant segment set is given by the user at the beginning of the segmentation process, and the algorithm stops when this set emerges.

4 Result Discussion

The results of the waterball segmentation method are demonstrated in Fig. 2 and Fig. 3 for a series of test images. The parameters of the algorithm were chosen individually in each case to achieve satisfactory performance and meaningful output.

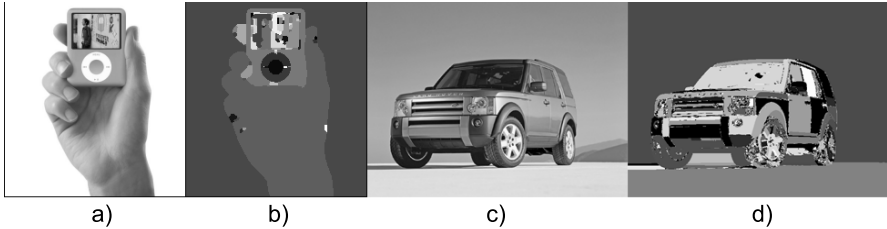


Fig. 2 Example images: hand (a), car (c), and the result of „waterball” segmentation of these images: (b), (d)

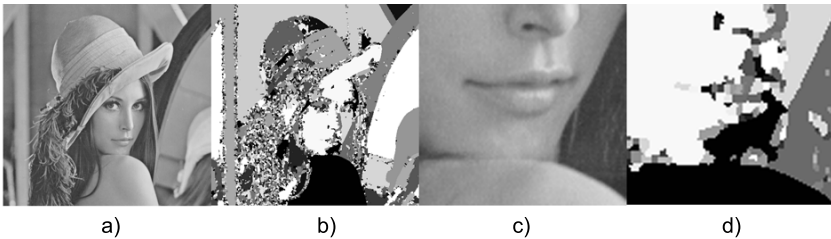


Fig. 3 Reference image „Lena”: original image (a), segmentation by „waterball” (b), detail of the original image(c), detail of the segmented image (d)

In Fig. 4 several iterations of the „waterball” method are shown for a chosen image. In this case, the best results are obtained with 13 iterations. It is worth noticing that if the dominant segment set parameters were chosen poorly and the iterations are allowed to continue beyond the desired, intuitive stop point, the fine details and smaller internal objects are gradually lost but the segmentation still produces viable macro-scale results for several more iterations. This effect is demonstrated in Fig. 4(e).

The „waterball” method is less effective in segmentation of images, where main objects of interest have weakly defined or blurred edges. During the edge enhancement step, these edges will disappear faster than desired and some internal segments laying close to these edges might be merged with the background. Another undesired effect is the presence of many small segments in the proximity of such weakly

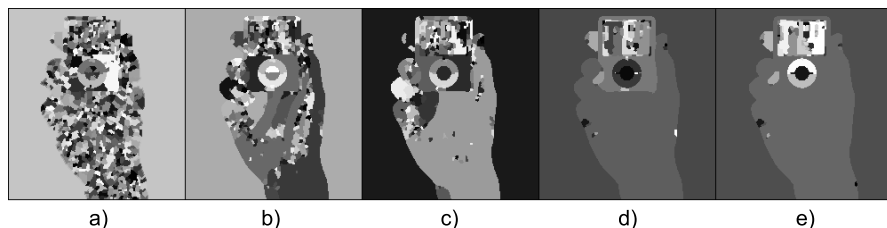


Fig. 4 Iterations of „waterball” algorithm for image „hand”: 1st (a), 3rd (b), 6th (c), 13th (d) and 15th (e)

defined edges, as can be observed in Fig. 3(d). The overall results, however, are on par with results reported in [3, 9] when considering the effectiveness of macro-scale object extraction.

The structure of the „waterball” algorithm allows for a wide degree of parallelization of calculations. One possible approach is to obtain several input images corresponding to several iterations of the edge enhancement steps, and then calculate watersheds for each of these images in parallel, on different processing nodes. Afterwards, the results can be evaluated using the dominant segment criterion and the best result can be chosen. This approach would allow significant shortening of the total calculation time, however, at the expense of increased memory consumption (as all the data structures associated with each input image would have to be stored in memory at the same time).

5 Conclusion

The „waterball” algorithm presented in this paper provides an efficient tool for segmentation of digital greyscale images and a viable alternative to existing methods based on classical watersheds and region merging. The oversegmentation was significantly reduced while the edges of macro-scale objects were faithfully reproduced. The method performs best when the objects of interest in the image are well defined, with strong edges and uniform background and in such cases, can be effectively used as a replacement of classical image segmentation algorithms.

References

- [1] Beucher, S.: The watershed transformation applied to image segmentation. *Scanning Microscopy International* 6, 299–314 (1991)
- [2] Bieniek, A., Moga, A.: An efficient watershed algorithm based on connected components. *Pattern Recognition* 33(6), 907–916 (2000)
- [3] Haris, K., Efstratiadis, S., Maglaveras, N., Katsaggelos, A.: Hybrid Image Segmentation Using Watersheds and Fast Region Merging. *IEEE Transactions on image processing* 7(12) (1998)

- [4] Lin, Y.C., Tsai, Y.P., Hung, Y.P., Shih, Z.C.: Comparison Between Immersion-Based and Toboggan- Based Watershed Image Segmentation. *IEEE Transactions on Image Processing* 15(3), 632–640 (2006)
- [5] Osma-Ruiz, V., Godino-Llorente, J.I., Saenz-Lechon, N., Gomez-Vilda, P.: An improved watershed algorithm based on efficient computation of shortest paths. *Pattern Recognition* 40(3), 1078–1090 (2007)
- [6] Rambabu, C., Chakrabarti, I.: An efficient immersion-based watershed transform method and its prototype architecture. *Journal of Systems Architecture* 53(4), 210–226 (2007)
- [7] Soille, P.: *Morphological Image Analysis - principles and applications*. Springer, Telos (1999)
- [8] Stoev, S.L.: RaFSi - A Fast Watershed Algorithm Based on Rainfalling Simulation. In: 8th International Conference on Computer Graphics, Visualization, and Interactive Digital Media (WSCG 2000), pp. 100–107 (2000)
- [9] Sun, H., Yang, J., Ren, M.: A fast watershed algorithm based on chain code and application in image segmentation. *Pattern Recognition Letters* 26(9), 1266–1274 (2005)
- [10] Świercz, M., Iwanowski, M.: Fast, Parallel Watershed Algorithm Based on Path Tracing. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) *IC-CVG 2010*. LNCS, vol. 6375, pp. 317–324. Springer, Heidelberg (2010)
- [11] Vincent, L.: Morphological Algorithms. In: Dougherty, E. (ed.) *Mathematical Morphology in Image Processing*, ch. 8, pp. 255–288. Marcel-Dekker, New York (1992)
- [12] Vincent, L., Soille, P.: Watersheds in digital spaces - an efficient solution based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(6), 583–598 (1991)
- [13] Wang, J., Lu, H., Eude, W., Liu, Q.: A Fast Region Merging Algorithm For Watershed Segmentation. In: 7th International Conference on Signal Processing (ICSP 2004), Beijing, China, vol. 1, pp. 781–784 (2004)

Kernel Postprocessing of Multispectral Images

Marcin Michalak and Adam Świtoński

Abstract. Multispectral analysis is the one of possible ways of skin disease detection. This short paper describes the nonparametrical way of multispectral image postprocessing that improves the quality of obtained pictures. The method below may be described as the regressional approach because it uses kernel regression function estimator as its essence. The algorithm called *HASKE* was developed as the time series predictor. Its simplification may be used for the postprocessing of multispectral images.

Keywords: Multispectral images analysis, nonparametrical regression, machine learning, *HASKE*.

1 Introduction

Every color may be described as the finite set of real values. Depending of the features we choose we obtain models like RGB, HSV and many others. When the number of color features, called channels, increases we say about multispectral or hyperspectral color definitions. We assume, that increasing the number of channels will give us more information about every pixel, or in other words the discernibility of different pixels will be higher.

This article describes the kernel method of multispectral image postprocessing and points the direction of its development to obtain the scalable postprocessing algorithm that would give also high quality of results.

Marcin Michalak

Central Mining Institute, Plac Gwarkow 1, 40-166 Katowice, Poland
e-mail: Marcin.Michalak@gig.eu

Marcin Michalak · Adam Świtoński

Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland
e-mail: {Marcin.Michalak, Adam.Switonski}@polsl.pl

Adam Świtoński

Polish-Japanese Institute of Information Technology,
ul. Koszykowa 86, 02-008 Warszawa, Poland

It is common, that measured samples (spectra) of the same color may be different due to the distance from the camera or the difference in lightening density. If the reference point is given, for example correct spectra of finite number of colors, it should be possible to describe the dependence between the observed spectrum and the expected one.

Next chapter introduces the research context of the presented paper. Then some theoretical aspects are described, especially the definition of *HASKE* algorithm that became the essential part of the kernel postprocessing. Afterwards experiments and their results are described. The paper ends with some conclusions and possible perspectives of further works.

2 Previous Research

Multispectral analysis becomes more popular as a tool for skin disaes diagnosis [6], [7]. In the paper [8] similar problem is raised: multispectral images for skin tumour tissue detection. On the base of acquired skin images and tumor regions pointed by medical experts some popular classifiers were build. Simultaneously, the other point of interest was defined: how does the measuring device influence obtained images? It occured, that for the well define given color (color with its spectrum) a wide range of spectra were obtained from the device (device is shown on the Fig. 1).

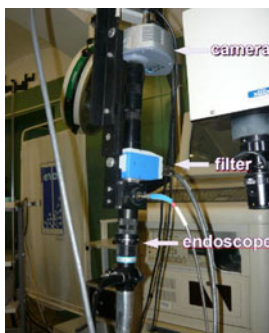


Fig. 1 Acquiring device.

In the paper [4] two approaches of multispectral images postprocessing were proposed. First of them (called classificational) give quite good results but has the disadvantage of being not scalable for colors that did not take the participation in building the classifier. Second one (called regressional) is scalable, but does not give satisfying results. This paper shows the results of regressional approach improvement.



Fig. 2 Colors from the ColorChecker.

3 Regressional Postprocessing

For the further reading of the paper the following notions are defined:

specimen - the real spectrum of the colour from the GretagMacbeth ColorChecker;
 sample - the measured spectrum from the ColorChecker;
 (color) profile - the aggregated information about all samples from the same colour of the ColorChecker.

The regressional approach of multispectral images postprocessing is based on the assumption that there exists some unknown regression function between the observed spectrum and the real one (close to the specimen). This dependence may be the one of the following:

- "1 vs 1": describes how the c -th color of the specimen depends on the c -th color of original spectrum;
- "1 vs all": describes how the c -th color of the specimen depend on the all image channels;
- "1 vs all ext.": extension of previous model wit the usage of additional input variables like the spectrum intergal.

3.1 Nadaraya–Watson Estimator

The Nadaraya–Watson estimator [5][10] is defined in the following way:

$$\tilde{f}(x) = \frac{\sum_{i=1}^n y_i K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)} \quad (1)$$

where $\tilde{f}(x)$ means the estimator of the $f(x)$ value, n is a number of training pairs (x, y) , K is a kernel function and h is the smoothing parameter.

As it is described in [7] and [9], the selection of h is more important than the selection of the kernel function. It may ocure that small values of the h cause that the estimator fits data too much. From the other hand big values of this parameter h lead to the estimator that oversmooths dependencies in the analysed set.

One of the most popular method of evaluation the h parametr is the analysis of the approximation of the *Mean Integrated Squared root Error (MISE)*.

$$MISE(h) = \int E \{ \tilde{f}_h(x) - f(x) \}^2 dx \tag{2}$$

The generalization of the above formula, with the introduction of *Integrated Squared Bias(ISB)* and *Integrated Variance (IV)*, leads to the *Asymptotic Mean Integrated Squared root Error (AMISE)*:

$$AMISE(h) = \frac{R(K)}{nh} + \frac{1}{4} \sigma_K^4 h^4 R(f'') \tag{3}$$

where $R(L) = \int L^2(x)dx$ and $\sigma_K^\alpha = \int_{-\infty}^\infty x^\alpha K(x)dx$. Optimization the *AMISE* in respect of h gives:

$$h_0 = \left[\frac{R(K)}{\sigma_K^4 R(f'')} \right]^{\frac{1}{5}} \cdot n^{-\frac{1}{5}} \tag{4}$$

The value of the expression $R(K)$ depends on the choosen kernel function K , but the value of the expression $R(f'')$ is unknown, so it is being replaced by some estimators. Further generalization leads to the following expression:

$$h_0 = 1,06 \min(\tilde{\sigma}, \tilde{R}/1, 34)n^{-1/5} \tag{5}$$

Details of derivations can be found in [7]. More advanced methods of h evaluation can be found in [9].

One of the most popular is the Epanechnikov kernel [2]. If V_q denotes the volume of the q -dimensional unit sphere in the \mathbb{R}^q space:

$$V_q = \begin{cases} \frac{(2\pi)^{q/2}}{2 \cdot 4 \cdot \dots \cdot n} & \text{for } q \text{ even} \\ \frac{2(2\pi)^{(q-1)/2}}{1 \cdot 3 \cdot \dots \cdot n} & \text{for } q \text{ odd} \end{cases}$$

then the Epanechnikov kernel is defined as:

$$K(x) = \frac{q+2}{2V_q} (1 - ||x||), ||x|| \leq 1$$

Other popular kernel function are listed in the Table 1 where $I(A)$ means the indicator of the set A .

3.2 HASKE Algorithm

Heuristic Adaptive Smoothing parameter Kernel Estimator (HASKE) [3] was developed as the kernel time series predictor. As the first step of the algorithm the mapping between the time series domain (t, x_t) and the regression domain (x_t, x_{t+p})

Table 1 Popular kernel functions.

kernel type	onedimensional	multidimensional
Epanechnikov	$K(x) = \frac{3}{4}(1-x^2)I(-1 < x < 1)$	$K(x) = \frac{q+2}{2V_q}(1- x)I(x < 1)$
Uniform	$K(x) = \frac{1}{2}I(-1 < x < 1)$	$K(x) = \frac{1}{V_q}I(x < 1)$
Triangular	$K(x) = (1- x)I(-1 < x < 1)$	$K(x) = \frac{q+1}{V_q}(1-\sqrt{ x })I(x < 1)$
Biweight	$K(x) = \frac{15}{16}(1-u^2)I(-1 < x < 1)$	$K(x) = \frac{(q+2)(q+4)}{V_q}(1- x)^2I(x < 1)$
Gaussian	$K(x) = \frac{1}{\sqrt{2\pi}}\exp -u^2/2$	$K(x) = (2\pi)^{-q/2}\exp(- x /2)$

is performed. The p parameter is usually a period of the time series. After the mapping, the final prediction is proceeded as the estimation of the regression function task that is divided into two steps.

Before the first step the division of the training set into the smaller training step and the tuning set is performed. Then the value of the smoothing parameter h , calculated on the smaller training set with the Eq. 5 is modified with the usage of parameter μ : $h' = \mu h$. The value of the μ parameter that gives the smallest error on the tune set is considered as the optimal one [1]. To avoid the underestimation (or overestimation) of the result the second *HASKE* parameter is introduced. The underestimation α_i is defined as the quotient of estimated value \tilde{y}_i and the real value y_i . Then, the final underestimation α is defined as the median of tune set underestimations values calculated with the usage of h' smoothing parameter and the Nadaraya–Watson estimator.

After the first step, when two new parameters μ and α are calculated, the final regression step is performed: on the basis of the (bigger) training set smoothing parameter is calculated, then its value is multiplied by the μ value and the result of Nadaraya–Watson estimator is divided by α .

For the purpose of using *HASKE* for simple regression there is no need to transform data from one space to another. Only division the train set into train and tune set must be performed. On the basis of μ and α the final formula of regression takes the following form:

$$\tilde{f}(x_t) = \frac{1}{\alpha} \frac{\sum_{i=1}^n y_i K\left(\frac{x_i - x_t}{h \cdot \mu}\right)}{\sum_{i=1}^n K\left(\frac{x_i - x_t}{h \cdot \mu}\right)} \tag{6}$$

4 Experiments and Results

The data set contained 31 456 observed spectra from 24 colors. Four of them were represented by 2130 instances, one was represented by 1238 instances and the rest

¹ For the time series it is a little more complicated and the notion of time series phase is introduced. Details are described in detail in [3].

of them were represented by 1142 instances. Specimen spectra were 36 channel and were interpolated linearly into 21 channel (domain of the acquisition device spectra). The data set was divided randomly into train set (28 311 objects) and the test set (3 145 objects).

As the quality measure the root of the mean value of the squared absolute error (*RMSAE*) was used:

$$RMSAE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{\tilde{c}_i - c_i}{c_i} \right)^2} \quad (7)$$

All three models of data dependence were examined: "1 vs 1", "1 vs all", "1 vs all ext.". For each of the model three regressors were used: Nadaraya–Watson (*NW*), *HASKE* without underestimation step (μ *HASKE*) and full *HASKE*. The result comparison is shown in the Table 2.

Table 2 Postprocessing results.

model	regressor		
	NW	μ <i>HASKE</i>	<i>HASKE</i>
1 vs 1	102,8%	93,2%	93,24%
1 vs all	100,0%	130,2%	236,4%
1 vs all (ext.)	100,0%	46,7%	42,3%

We may see that the usage of *HASKE* gives different results for different models of regressional dependency. The best results may be observed for the third one. In this case it is also visible that the following steps of *HASKE* give better results than the typical regression with Nadaraya–Watson estimator.

5 Conclusions and Further Works

Referring to the results of the previous research [4] in this paper the regressional approach of multispectral images postprocessing was developed. The usage of time series dedicated *HASKE* algorithm gave better results than non-modified Nadaraya–Watson kernel estimator. The adaptive modification of the smoothing parameter value that is the essential part of *HASKE* should be regarded as the main reason of the results improvement. This means that further works should focus on the problem of *h* parameter evaluation.

It also occurred that adaptive methods of smoothing parameter evaluation should be applied for the third model of spectra dependencies representation. It may be also interesting to expand this model with several new variables like the typical statistics: mean, maximal and minimal value, standard deviation etc.

It may be also interesting to recognize which spectra channels provide most of the color information. In other words – which channels are the most useless ones and make the models more complicated.

Acknowledgments. This work was financed from the Polish Ministry of Science and Higher Education resources in 2009-2012 years as a research project.

References

- [1] Blum, A., Zalaudek, I., Argenziano, G.: Digital Image Analysis for Diagnosis of Skin Tumors. *Seminars in Cutaneous Medicine and Surgery* 27(1), 11–15 (2008)
- [2] Epanechnikov, V.A.: Nonparametric Estimation of a Multivariate Probability Density. *Theory of Probab. and its Appl.* 14, 153–158 (1969)
- [3] Michalak, M.: Time series prediction using new adaptive kernel estimators. *Adv. in Intell. and Soft Comput.* 57, 229–236 (2009)
- [4] Michalak, M., Świtoński, A.: Spectrum evaluation on multispectral images by machine learning techniques. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) *ICCVG 2010. LNCS*, vol. 6375, pp. 126–133. Springer, Heidelberg (2010)
- [5] Nadaraya, E.A.: On estimating regression. *Theory of Probab. and its Appl.* 9, 141–142 (1964)
- [6] Prigent, S., Descombes, X., Zugaj, D., Martel, P., Zerubia, J.: Multi-spectral image analysis for skin pigmentation classification. In: *Proc. of IEEE Int. Conf. on Image Process (ICIP)*, pp. 3641–3644 (2010)
- [7] Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, Boca Raton (1986)
- [8] Świtoński, A., Michalak, M., Josiński, H., Wojciechowski, K.: Detection of tumor tissue based on the multispectral imaging. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) *ICCVG 2010. LNCS*, vol. 6375, pp. 325–333. Springer, Heidelberg (2010)
- [9] Turlach, B.A.: *Bandwidth Selection in Kernel Density Estimation: A Review*. C.O.R.E. and Institut de Statistique, Universite Catholique de Louvain (1993)
- [10] Watson, G.S.: Smooth Regression Analysis. *Sankhya - The Indian J. of Stat.* 26, 359–372 (1964)

A Semi-local Topological Constraint for Efficient Detection of Near-Duplicate Image Fragments

Mariusz Paradowski and Andrzej Śluzek

Abstract. Image matching methods are an important branch of computer vision and have many possible applications, i.e. robotics, navigation, etc. Their goal is to detect near-duplicate images, sub-images or even localize image fragments. The paper addresses the last, and the most difficult problem: simultaneous localization of multiple image fragments in images of unknown content. A complete lack of any *a priori* knowledge is assumed, including no information of the number of fragments. The presented approach combines low level vision techniques and high level spatial constraints. Photometric properties of the images are represented using key-regions. Spatial consistency is verified using the topology of images. A graph of topologically consistent key-regions is created. It allows efficient localization of entire near-duplicate image fragments.

1 Introduction

Local approaches, where the data representing visual and spatial properties of images are *key-regions* (also referred to as *keypoints*), are typical tools for near-duplicate detection in image matching. In most cases, key-regions are elliptical areas within the image, which are extracted using various key-region *detectors*, e.g. in places of high gradient changes (Harris-Affine detector [5]). Such key-regions are characterized by feature vectors typically built by key-region *descriptors*, e.g. *SIFT*, [4], is one of the most popular choices. The methods of key-region detection and description are designed to be invariant under typical geometric and photometric distortions, which makes key-region matching a feasible method of identifying visually similar image patches.

However, even the most reliable approaches usually match correctly only a fraction (sometimes very small) of the returned key-region pairs. To remove incorrect

Mariusz Paradowski · Andrzej Śluzek

Institute of Informatics, Wrocław University of Technology

School of Computer Engineering, Nanyang Technological University

Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University

matches (outliers), additional steps are needed. Various kinds of spatial models are used to verify the correctness of key-region matching. These models are based on either image geometry or topology. They may be *unimodal* (matching an image against a single model or a sub-image) or *multimodal* (to find multiple similar fragments in matched images). In this paper we discuss a multimodal approach based on image topology.

Problem formulation

In this work, we generally assume no *a priori* knowledge about the processed images, i.e. the method is designed to handle random visual contents so that it is not possible to build any model of images or to apply any typical machine learning approaches. Therefore, we define the objective in a very general and straightforward manner: *Given two random images, find all near-duplicate image fragments which appeared on these two images.*

Selected prior works

As mentioned above, the correctness of key-region matching is verified using either *unimodal* and *multimodal* approaches (the former being more common and popular). In the unimodal techniques, the objective is to determine the optimum image transformation which maximizes the correspondences between the actual and model-mapped coordinates of matched key-regions. A range of approaches is possible, very often based on a popular RANSAC algorithm [1]. Affine [3, 4], homography [2] or non-linear [12] transformations are used to model the image mappings. *Multimodal* approaches (though discussed in the research literature) are less popular. They are designed to handle a more difficult problem, i.e. simultaneous detection and localization of multiple near-duplicate fragments in matched images. Similarity [13] and affine [7, 9] transformations have been used to build geometry-based constraints verifying the correctness of key-region matching.

In the presented work, we focus on image topology instead, i.e. local topological constraints are applied to verify whether groups of matched key-regions actually form similar fragments, [8]. This is obviously more general compared to the geometrical approach. However, such a generality may result in a lower precision of the topological methods.

The idea of semi-local topological pseudo-invariants can be traced back to the works of Schmid and Mohr [10, 11]. They suggested the correctness of keypoint matching could be verified by processing neighborhoods of key-regions. A pair of keypoints is considered correctly matched, if a certain fraction (50% in [10]) of their nearest neighbors in both images are also matched. Filtered nearest neighbors of the key-region pair have to follow the similarity transformation, which is less generic than the affine transformation. In our earlier research, we have generalized this approach and proposed another topological constraint based on the orientation of vectors [8], which puts much less assumptions on the geometry of the neighborhood. However, computational costs are higher because of its complexity which is

equal to $O(nN^3)$, where n is the number of keypoint pairs and N is the size of the neighborhood. Thus, the neighborhood N has to be very small in practice.

We have verified the usefulness of the topologically consistent neighborhoods (as suggested in Schmid–Mohr approach) in our prior work [6]. Similarly to the original report [10], we have used the semi-local invariant to calculate distances between many images, i.e. to pre-retrieve images possibly containing fragments similar to the query. In this paper, we apply the same constraint to a much more difficult task – to accurately locate multiple similar fragments existing in the matched images. We have addressed different aspects of such a problem in our previous researches. In [9], we proposed a geometrical, triangle based method of relatively high computational complexity ($O(nN^2)$). Its faster ($O(nN)$) complexity variant with more assumptions and lower quality has been presented in [7]. A topological method based on vector orientation (of $O(nN^3)$ complexity) has been reported in [8]. Our objective is, therefore, to develop a method providing similar (or better) performances at lower computational costs.

2 The Proposed Method

The main contribution of this paper is a method capable to detect (without any image model and/or learning) multiple near-duplicate fragments in matched images at $O(nN)$ complexity. The proposed method combines two ideas: the neighborhood-based topological constraint [6, 10] used to verify if a pair of key-regions is reliably matched, and the topological graph [8] used to merge key-regions into image fragments. The general concept of the method can be represented by the following pseudocode:

```

1: Detect key-regions in each image (off-line).
2: Find N neighborhoods for each key-region (off-line).
3: Match key-regions between two images  $\mathcal{I}$  and  $\mathcal{J}$ .
4: for all key region pairs  $R(\mathcal{I}, \mathcal{J})$  do
5:   Identify all matched pairs of key-regions within the neighborhoods for each matched key-region pair  $(r_{\mathcal{I}}, r_{\mathcal{J}})$ .
6:   Count the number of identified key-region pairs in both images
7:   if number of key-region pairs higher than threshold in both images then
8:     Mark the  $(r_{\mathcal{I}}, r_{\mathcal{J}})$  key-region pair as correct
9:     Connect key-regions  $r_{\mathcal{I}}$  and  $r_{\mathcal{J}}$  to their matched neighbors
10:  end if
11: end for
12: Group all correct regions according to connections.
13: Calculate convex hull for each group.

```

2.1 Local Topological Constraint

We assume that if a pair of key-regions is correctly matched, more matches should be consistently expected in the neighborhoods of both key-regions. Such a *consistency of the neighborhoods* is verified by counting matched key-region pairs within both neighborhoods. In other words, a single evidence that two regions are visually similar may not be enough. However, if we have multiple such evidences, the regions should be considered visually similar. This idea is illustrated in Fig. 1.

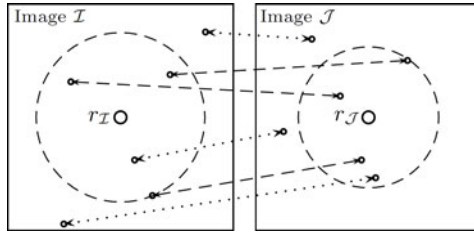


Fig. 1 Illustration of the neighborhood-based topological constraint. Only these key-region pairs are correct which are surrounded by a sufficient number of matching pairs of other key-regions.

Given an Euclidean distance d , the spatial neighbourhood $\mathcal{N}(r_{\mathcal{X}})$ of a size s for a key-region $r_{\mathcal{X}}$ from an image (set of key-regions) \mathcal{X} is equal to:

$$\mathcal{N}(r_{\mathcal{X}}) = \arg \min_{N \in \mathcal{X}} \sum_{n \in N} d(n, r_{\mathcal{X}}), \quad |N| = s. \quad (1)$$

Given two images \mathcal{I}, \mathcal{J} , a pair of key-regions $(r_{\mathcal{I}}, r_{\mathcal{J}}) \in R(\mathcal{I}, \mathcal{J})$ and a threshold t , the neighborhood-based topological constraint $\mathcal{T}(\mathcal{I}, \mathcal{J}, r_{\mathcal{I}}, r_{\mathcal{J}}, t)$ is defined as:

$$\mathcal{T}(\mathcal{I}, \mathcal{J}, r_{\mathcal{I}}, r_{\mathcal{J}}, t) = \left| [\mathcal{N}(r_{\mathcal{I}}) \times \mathcal{N}(r_{\mathcal{J}})] \cap R(\mathcal{I}, \mathcal{J}) \right| > t, \quad (2)$$

where $R(\mathcal{I}, \mathcal{J})$ is a set of key-region pairs, \mathcal{N} is the spatial neighborhood function, and \times stands for Cartesian product.

The idea of multiple-evidence key-region pairs does have an important advantage: it is very simple to verify. Given key-region neighborhoods, we only need to check how many matched pairs belong to the neighborhoods. For a single pair of key-regions, this operation takes only $O(N)$ so that the total complexity for all matched pairs of key-regions is equal to $O(nN)$.

To demonstrate that such a topological constraint has the ability to filter out only correct keypoint matches, a number of examples is shown in Fig. 2. In general, the majority of filtered keypoint pairs (small circles) are correct.

2.2 Topological Graph

Multiple evidences supporting similarity of image fragments can be conveniently represented by a graph (which will be referred to as *topological graph* [8]). Each node of the topological graph corresponds to a pair of matched key-regions. If two pairs of matched key-regions are found evidences supporting similarity of an image fragment (as explained above), and edge is created between the corresponding nodes of the graph.



(a) Despite background clutter, small road signs are filtered correctly. (b) The method is invariant to typical image transformations. (c) Weakness of the approach – matches are correct, but the region is incorrectly detected.

Fig. 2 Image pairs with encircled filtered keypoints.

Now, the retrieval of similar image fragments can be considered the connected component search. In fact, the routine is a classic 1-connected component search algorithm, which takes only $O(nN)$, which is the maximum possible number of edges in the graph. Exemplary components found within the topological graph are shown in Fig. 3. Key-regions are marked with small circles and lines represent connections between them. Because of a high acceptance threshold (the percentage of key-region matches with a neighborhood) the number of graph edges, i.e. the number of lines in the figure, is large.

Once the topological graph is constructed and its nodes are grouped into connected components, we create convex hulls of such groups of (separately in both images). The convex hulls represent the estimated outlines of similar image fragments.

3 Experimental Verification

The experimental verification of the method has been performed on a database containing 9900 image pairs (<http://www.ii.pwr.wroc.pl/~visible/>). About 15% of the pairs share at least one common fragment, many of these pairs contain multiple common fragments. This makes the data set suitable both for testing *precision* (most pairs do not share similar content) and *recall* (a sufficiently large number of pairs share multiple objects). The database contains both indoor and outdoor scenes, with diversified objects. Background clutter is generally very high. The quality measurement is performed using classic *precision* and *recall* measures, in two aspects however. First, we only determine if the similar fragments are correctly extracted



(a) Objects are correctly separated. Only a single keypoint pair is incorrect. (b) The graph is dense, there are many evidences to support the similarity. (c) Occlusions do not pose problems, because there is no point pair is incorrect.

Fig. 3 Relations between keypoint pairs – the topological graph.

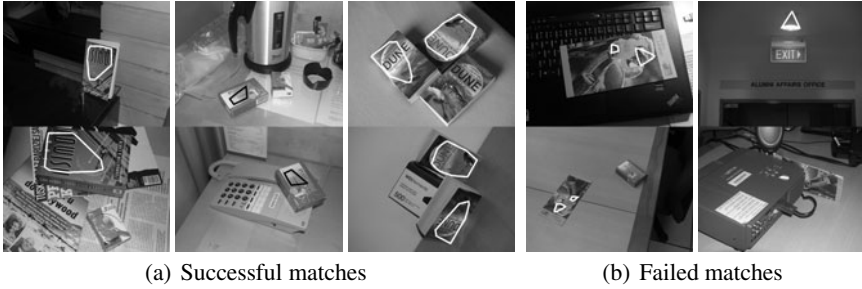
(regardless how well their shape are estimated). In the second aspect, we measure how accurately the areas of similar fragments are outlined.

The proposed matching method has two parameters – the neighborhood size N and the cutoff threshold t . Due to $O(nN)$ complexity the neighborhood size setting linearly impacts the method execution time. Thus, the parameter should not be excessively increased in order to avoid slowdowns. However, too small neighborhoods deteriorate *precision*. The cutoff threshold affects only the achieved quality of results. Increasing the cutoff threshold means that more evidences are required to support similarity. The results obtained for different parameter settings are presented in Tab. 1. The achieved *precision* is well above 90% both in object and area aspects. Exemplary similar fragments found within the database images are presented in Fig. 4. In fact, the algorithm’s ability to effectively process large neighborhoods (e.g. $N = 140$) is one of the advantages of this approach. Our previous approach (the orientation-based topological method [8]) has been limited only to very small neighborhoods (recommended $N = 30$) due to $O(nN^3)$ complexity. We have also tested our approach on the threshold setup suggested in [10] ($t = 0.5N$). Almost all key-regions were filtered out and the achieved quality was low.

To summarize the experimental verification of the method, we compare it with other image fragment matching approaches. All of these approaches have higher computational complexity. Quality of results achieved by the new method is fully satisfying. It achieves very high precision – over 90%, while keeping recall at competing levels. The final comparison is shown in Tab. 2.

Table 1 Average quality results for 9900 image pairs.

		neighborhood size			
		80	100	120	140
object based measurement (precision/recall/f-score)					
threshold	12%	0.69/0.85/0.76	0.75/0.85/0.80	0.84/0.83/0.84	0.86/0.83/0.85
	14%	0.74/0.83/0.78	0.83/0.81/0.82	0.89/0.79/0.84	0.92/0.78/0.84
	16%	0.82/0.78/0.80	0.88/0.78/0.83	0.90/0.76/0.82	0.93/0.75/0.83
	20%	0.85/0.71/0.77	0.88/0.68/0.77	0.91/0.65/0.76	0.94/0.63/0.75
area based measurement (precision/recall/f-score)					
threshold	12%	0.75/0.46/0.57	0.79/0.49/0.61	0.86/0.49/0.63	0.87/0.52/0.65
	14%	0.81/0.43/0.56	0.88/0.44/0.59	0.91/0.44/0.60	0.93/0.45/0.60
	16%	0.88/0.37/0.52	0.92/0.39/0.55	0.93/0.37/0.53	0.94/0.38/0.54
	20%	0.92/0.28/0.43	0.93/0.28/0.43	0.94/0.28/0.43	0.96/0.28/0.43

**Fig. 4** Matched image fragments using the neighborhood based topological constraint and the topological graph.**Table 2** Comparison of image fragment matching methods.

<i>image fragment matching method</i>	<i>fragment matching</i>	<i>key regions matching</i>	<i>prec. [obj.]</i>	<i>recall [obj.]</i>	<i>prec. [area]</i>	<i>recall [area]</i>
topological-angles [8]	$O(nN^3)$	$O(kn^2)$	0.94	0.93	0.63	0.79
geometrical-triangles [9]	$O(n^2 + nN^2)$	$O(kn^2)$	0.96	0.82	0.95	0.65
geometrical-ellipses [7]	$O(n^2 + nN)$	$O(kn^2)$	0.95	0.65	0.87	0.49
topological-neighbors	$O(nN)$	$O(kn^2)$	0.92	0.78	0.93	0.45

4 Summary

A novel method for matching image fragments has been presented. It is based on a local topological constraint defined over the spatial neighborhoods of key-region

pairs. The main advantage of the method is its low computational complexity equal to $O(nN)$, where n is the number of key-region pairs and N is the size of the spatial neighborhood. The method is able to compete with much more complex approaches. On the test data set we achieved up to 92% precision for the object-based aspect and 93% for the area aspect of measurements. These results are comparable with the geometrical approach, which has much higher computational complexity.

Acknowledgement. This work is partially financed from the Ministry of Science and Higher Education Republic of Poland resources in 2008–2010 years as a Poland–Singapore joint research project 65/N-SINGAPORE/2007/0.

References

- [1] Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: Buxton, B.F., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1065, pp. 683–695. Springer, Heidelberg (1996)
- [2] Kannala, J., Salo, M., Heikkila, J.: Algorithms for computing a planar homography from conics in correspondence (2006)
- [3] Lowe, D.G.: Object recognition from local scale-invariant features. In: Proc. 7th IEEE Int. Conf. Computer Vision, vol. 2, pp. 1150–1157 (1999)
- [4] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
- [5] Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. *International Journal of Computer Vision* 60, 63–86 (2004)
- [6] Paradowski, M., Śluzek, A.: Automatic visual object formation using image fragment matching. In: Proc. International Multiconference on Computer Science and Information Technology, pp. 97–104 (2010)
- [7] Paradowski, M., Śluzek, A.: Detection of image fragments related by affine transforms: Matching triangles and ellipses. In: Proc. International Conference on Information Science and Applications, vol. 1, pp. 189–196 (2010)
- [8] Paradowski, M., Śluzek, A.: Keypoint-based detection of near-duplicate image fragments using image geometry and topology. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010. LNCS, vol. 6375, pp. 175–182. Springer, Heidelberg (2010)
- [9] Paradowski, M., Śluzek, A.: Matching planar fragments in images using histograms of decomposed affine transforms. In: 2nd review in TPAMI (2010)
- [10] Schmid, C., Mohr, R.: Object recognition using local characterization and semi-local constraints. Technical report, INRIA (1996)
- [11] Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. *IEEE TPAMI* 19(5), 530–535 (1997)
- [12] Yang, D., Sluzek, A.: A low-dimensional local descriptor incorporating tps warping for image matching. *Image and Vision Computing* (2009), doi:10.1016/j.imavis.2009.12.003
- [13] Zhao, W.-L., Ngo, C.-W.: Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection. *IEEE Trans. on Image Processing* 2, 412–423 (2009)

Morphology-Based Method for Reconstruction of Invisible Road Parts on Remote Sensing Imagery and Digitized Maps

Bartłomiej Zielinski and Marcin Iwanowski

Abstract. Road detection in remote sensing imagery is an important field of interest in remote scene analysis. In complex process of road network extraction one has to cope with many separate problems. One is presence of objects which are similar (or identical) in colour to road surface. Another is absence of some road fragments in the image. In this paper we propose a new approach with use of morphological image processing to address aforementioned issues. Moreover, we claim it works well for road detection on maps. Our experiments prove effectiveness of proposed solution. Along with relative simplicity, our proposal presents a convenient method for road network reconstruction.

1 Introduction

1.1 *The Background*

Increasing availability of satellite and aerial images contributed to growing interest in remote scene analysis. Demands emerging from different directions — industry, military, government agencies etc. — intensified involvement in remote sensing research which mainly focuses on detecting certain type of objects in examined image (or series of images). Multiple image analysis might be performed while discovering seasonal changes in the environment, updating geographic information systems, monitoring disasters like floods and conflagrations or determining their range of damage.

Among many subjects in remote sensing domain, road detection has been a topic of significant importance. Despite considerable research, no unailing solution has appeared. Numerous sub-problems arising in the course of extracting roads make

Bartłomiej Zielinski · Marcin Iwanowski
Institute of Control and Industrial Electronics,
Warsaw University of Technology,
ul.Koszykowa 75, 00-662 Warszawa Poland
e-mail: [{{zielinsb,iwanowski}@ee.pw.edu.pl](mailto:{zielinsb,iwanowski}@ee.pw.edu.pl)

the task not a trivial one. There are many potential ways of handling the question. Some of them offer ease of overcoming certain difficulties. However, there are some essential issues one has to deal with while designing a road detection tool.

First of all, some road fragments are invisible, as they lead underground or simply might be covered by treetops. It imposes necessity of guessing where some parts of a road are, because this information the initial image sometimes just lacks.

Secondly, some objects in an image might be of exactly same colour as roads. Sheds, buildings and a whole range of stuff similar to road surface make analysis even more onerous.

Thirdly, various types of roads appear in the image. Especially suburban areas are rich in roads of diverse surfaces: tarmac, concrete, dirt to name just a few. Every one of them is of a different colour which makes initial preselection pretty difficult task. The more so, because we cannot assume that all the images had been taken in exactly same lighting conditions.

By and large, even sophisticated methods of analyzing image need a solution of getting rid of unwanted objects as well as of merging roads into a network by filling gaps between potential road segments.

1.2 Scope of the Paper

This paper shows a mathematical morphology based approach solving separate problems which exist in road detection process. The approach handles with binary images, so a preprocessing step is required. It can be some complex preselection as well as simple thresholding method.

One of the problems the paper is addressing is presence of non-elongated objects in an image. Such objects might be suspected of being anything else but roads and generally can be safely removed. We solve the issue by erasing all the pixels which do not belong to lines longer than certain value.

The other problem is lack of some road fragments due to covering effects. Our algorithm merges near road parts and so reproduces road network.

1.3 Previous Works

Since the beginning of research in the area many different ideas have arisen. [2] presents a survey on road extraction methods. In [3] pattern matching is used to search for road parts. [6] offers an artificial neural network approach, while [8] proposes reconstructing roads with snakes and multiresolution analysis. In [1] a region competition algorithm is presented. [11] introduces road tracking with particle filtering. [10] and [7] apply Markov random fields and in [9] so called line segment matching is shown. [5] describes a road tracing method based on road footprints classification. Finally, in [4] road tracking by active testing is proposed.

2 Proposed Method

Mathematical morphology is a well known approach in image analysis. The main idea underlying its methods consists in taking advantage of so called structuring element which defines neighbourhood for every pixel of the image. Probing an image with structuring element of different shapes and by applying various morphological operators enables designing powerful algorithms for image processing purposes.

2.1 Preprocessing

As our solution operates with binary data, initial image has to be subjected with a preprocessing procedure. Presented method is meant to cope with noisy and insufficient information images. Performing simple thresholding of an image produces results which are definitely incomparable to outcomes of complex operations. However, our approach eliminates necessity of harnessing extremely sophisticated preprocessing methodology. Obviously, one may feel free to take any initial steps, as long as they output binary image ready for removing redundant objects and linking divided road fragments.

In our procedure we have used colour thresholding to preselect potential road segments. As a result, we have obtained images with separated road network and some unwanted objects as well.

2.2 Binary Morphological Filtering

Let $L(\alpha, l)$ be asymmetric and $\widehat{L}(\alpha, l)$ symmetric linear structuring elements specified as straight lines of angle α and length l . Examples of such structuring elements are shown in Fig. 1.

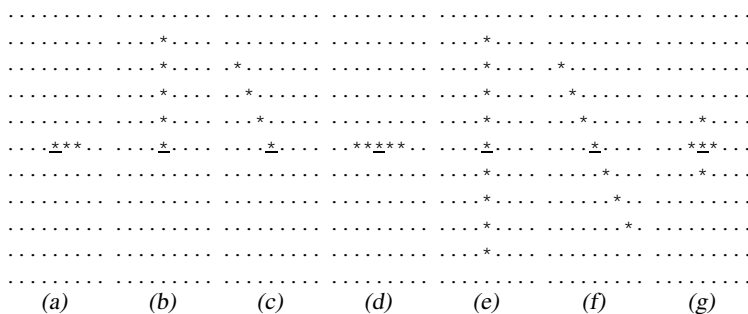


Fig. 1 Linear and elementary structuring elements: (a) $L(0, 3)$, (b) $L(\frac{1}{2}\pi, 5)$, (c) $L(\frac{3}{4}\pi, 4)$, (d) $\widehat{L}(0, 5)$, (e) $\widehat{L}(\frac{1}{2}\pi, 9)$, (f) $\widehat{L}(\frac{3}{4}\pi, 7)$, (g) elementary 4-connected \mathcal{N}_4 ; * stands for the pixel belonging to the structuring element, underlined pixel is its center.

Erosion and dilation operators are defined as follows, respectively:

$$F \ominus B = \bigcap_{b \in B} F_{[-b]}, \quad F \oplus B = \bigcup_{b \in B} F_{[-b]}, \quad (1)$$

where $F_{[-b]}$ stands for translation of image F by vector $-b$ and B stands for a structuring element.

Morphological filters of opening and closing are the combination of erosion and dilation performed successively:

$$F \circ B = (F \ominus B) \oplus B^T, \quad F \bullet B = (F \oplus B) \ominus B^T, \quad (2)$$

where \circ operator stands for opening, \bullet for closing and B^T is a transposed structuring element B (i.e. $B^T = \{-p : p \in B\}$).

Our goal is to define a filter that detects straight lines longer than a fixed threshold, at all possible angles. We achieve this by applying superposition of directional openings.

Basing on above, we introduce the following filter:

$$G = \bigcup_{\alpha \in A} [F \circ \widehat{L}(\alpha, l)], \quad (3)$$

where $A = \{s, 2s, 3s, \dots, \pi\}$ is a set of angles and s stands for a step value. Because of symmetry, $\widehat{L}(\alpha, l) = \widehat{L}(\alpha + \pi, l)$ and that is why set A contains angles from $\langle 0, \pi \rangle$.

The filter erases all the pixels in the image F except those which are part of segments of length of at least l , inclined at an angle $\alpha \in A$. By applying the filter with small value of step s between α angles, both accuracy and computational cost are increased. On the contrary, greater step s can save calculation time but the filter may skip some segments and therefore remove lines which should not have been deleted. Nevertheless, one should keep in mind that for little l values and even quite distant α values, structuring elements might be exactly the same.

2.3 Merging Split Road Parts

In order to detect invisible road fragments we propose an approach combining directional opening, dilation and closing.

Let us denote:

$$O_{\alpha, l} = F \circ \widehat{L}(\alpha, l), \quad (4)$$

$$D_{\alpha, l_1, l_2} = O_{\alpha, l_1} \oplus L(\alpha, l_2), \quad (5)$$

$$M_{\alpha, \beta, l_1, l_2} = \{[D_{\alpha, l_1, l_2} \cap D_{\beta, l_1, l_2}] \cup O_{\alpha, l_1}\} \bullet L(\alpha, l_2), \quad (6)$$

where F is initial binary image and $\alpha \in \langle 0, 2\pi \rangle$.

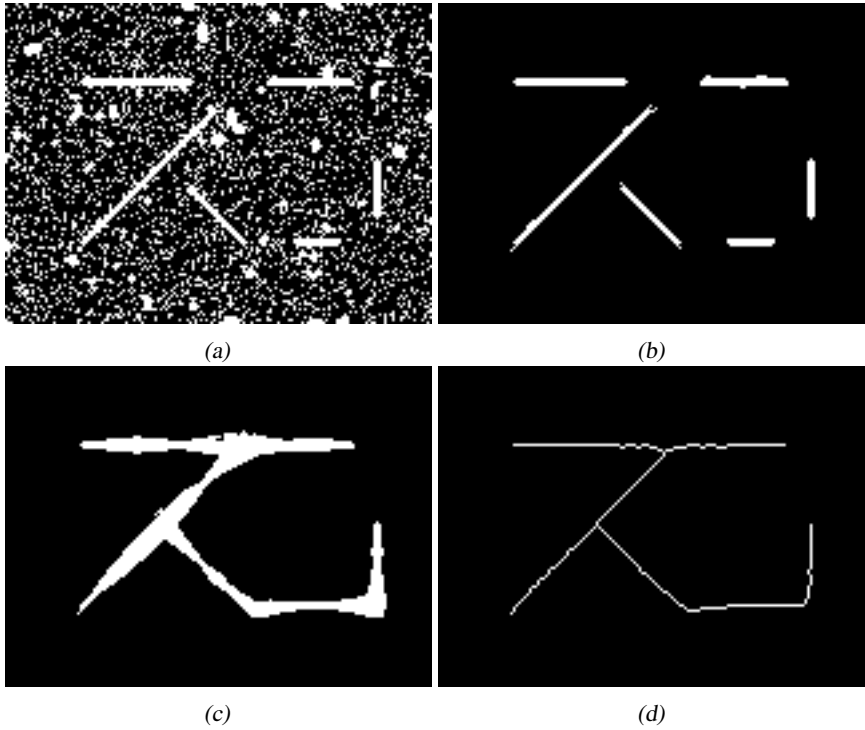


Fig. 2 Trial image: (a) noisy input image, (b) denoised image, (c) image with merged lines, (d) final road network.

Image I_{l_1, l_2} with merged lines is produced by performing:

$$I_{l_1, l_2} = \bigcup_{(\alpha, \beta)} [M_{\alpha, \beta, l_1, l_2} \cup M_{\beta, \alpha, l_1, l_2}], \quad (7)$$

such that (α, β) is a pair of any angles in $\langle 0, 2\pi \rangle$, satisfying the condition: $|\alpha - \beta| < \delta$, where δ stands for minimal angle difference.

To obtain the road network image I_{l_1, l_2} is smoothed by closing:

$$J_{l_1, l_2} = I_{l_1, l_2} \bullet \mathcal{N}_4 \quad (8)$$

and finally thinning objects to lines is accomplished. This eventual operation results in presenting roads as one pixel thick lines.

Minimal angle difference δ has been introduced in order to avoid lengthening of wide dead ends. With no δ condition, respectively broad objects would reshape, as intersection of dilated component lines may produce nonempty set. In fact, regardless of δ value, it might occur in case of wide roads, however thinning operation

ensures, at most, inconsiderable deflection, whereas in case of an elongated dead end, thinning operation is useless.

Let us investigate the algorithm on the following trial example (Fig. 2). Firstly, initial binary image (Fig. 2(a)) is repeatedly opened (Eq. 4) with different α angles. Union of result images (Eq. 3) is devoid of short lines, thus it lacks objects which most likely are not parts of a road (Fig. 2(b)). Next, pairs of images acquired by applying opening with different α values are dilated (Eq. 5) and their intersection is determined. Subsequently, union of previously fixed intersection with result of opening is closed (Eq. 6). Image with reconstructed road network (Fig. 2(c)) is an outcome of Eq. 7. Eventually, the image is smoothed by closing (Eq. 8) and thinned. Final result is presented in Fig. 2(d).

3 Results

The method has been tested on variety of images. We show possible application to recovering roads on remote sensed images and on maps.

Assuming the length of the longest gap between split road parts we want to merge is longer than the length of the longest unwanted object in the image, we choose proper l_1 and l_2 values. For all the images examined in this study we have applied minimal angle difference δ equal 30 degrees.

Initial colour images are presented as grayscale images.

3.1 Test Results on Images

We present two real life examples of images depicting exurbs. In both cases some parts of road surface are invisible, hidden behind trees. Moreover, images contain objects unable to remove by thresholding. With use of our method we managed to get rid of them and reconstruct road trace. In Fig. 3-8 consecutive image processing steps have been shown.



Fig. 3 Initial grayscale images.

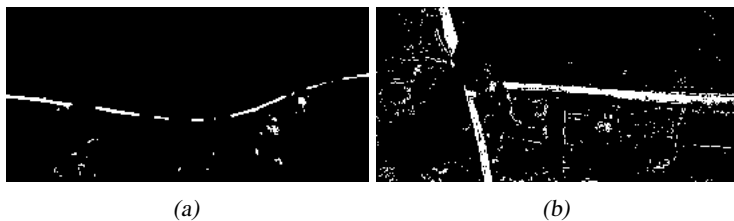


Fig. 4 Thresholding results.



Fig. 5 Images with non-elongated objects removed.



Fig. 6 Images with merged road segments.



Fig. 7 Images smoothed by closing.

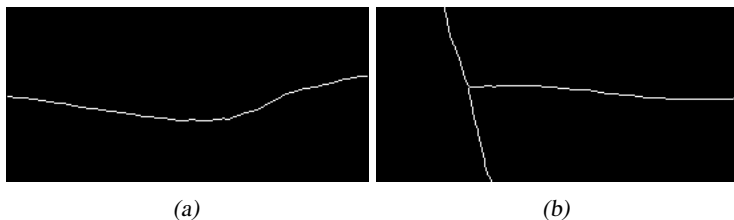


Fig. 8 Final images with roads thinned to lines.

3.2 Test Results on Maps

Our solution may also be useful in digitizing of archival maps. Maps are often covered with notations or marked with different signs and symbols. As a consequence, roads (or other important lines like borders, railways etc.) might not be entirely visible. As there is no problem with new maps which exist in digital form, in digitization of archival ones reconstruction measures are definitely required.

We present two images of maps. In both of them covering effects occur. In Fig. 9-14 we show consecutive processing steps leading to road reconstruction.

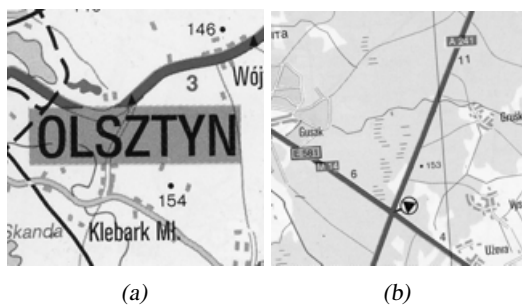


Fig. 9 Initial grayscale maps.

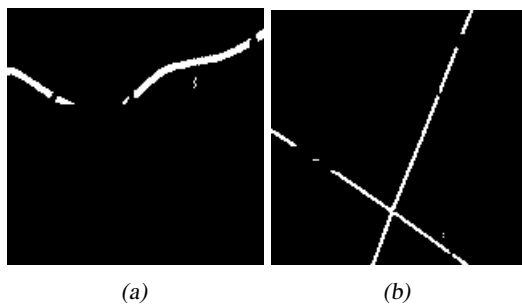


Fig. 10 Thresholding results.

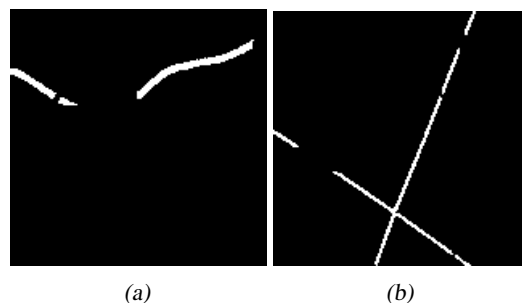


Fig. 11 Maps with non-elongated objects removed.

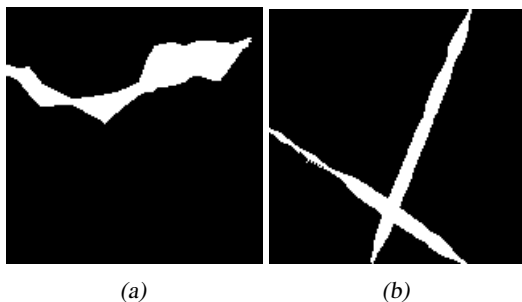


Fig. 12 Maps with merged road segments.

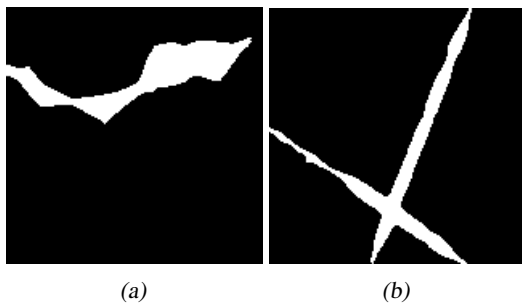


Fig. 13 Maps smoothed by closing.

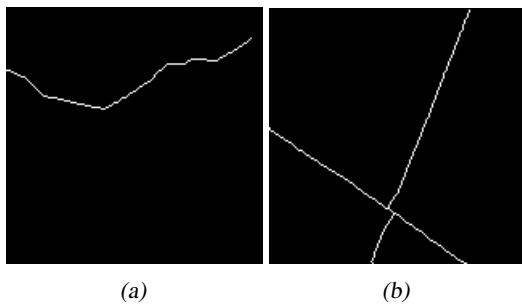


Fig. 14 Final maps with roads thinned to lines.

4 Conclusions

In this paper we have introduced a solution which might be useful as a component part of an algorithm detecting roads on satellite or aerial images and on maps. The approach is particularly serviceable while dealing with limited visibility of roads. Presented methodology is based upon mathematical morphology tools. Not only does it allow removing objects which — due to their shape — do not seem to be road parts, but enables merging separated road fragments as well.

Test results of both synthetic, real remote sensed images and maps confirm usefulness of proposed approach and justify its usage in road reconstruction problem.

References

- [1] Amo, M., Martinez, F., Torre, M.: Road Extraction From Aerial Images Using a Region Competition Algorithm. *IEEE Transactions on Image Processing* 15(5) (2006)
- [2] Auclair Fortier, M.F., Ziou, D., Armenakis, C., Wang, S.: Survey of Work on Road Extraction in Aerial and Satellite Images. Département de mathématiques et d'informatique, Université de Sherbrooke, Technical Report no. 247 (1999)
- [3] Bajcsy, R., Tavakoli, M.: Computer recognition of roads from satellite pictures. *IEEE Transactions on Systems, Man and Cybernetics* SMC-6(9) (1976)
- [4] Geman, D., Jedynak, B.: An Active Testing Model for Tracking Roads in Satellite Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(1) (1996)
- [5] Hu, J., Razdan, A., Femiani, J.C., Cui, M., Wonka, P.: Road Network Extraction and Intersection Detection from Aerial Images by Tracking Road Footprints. *IEEE Transactions on Geoscience and Remote Sensing* 45(12) (2007)
- [6] Mokhtarzade, M., Valadan Zoej, M.J.: Road detection from high-resolution satellite images using artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation* 9 (2007)
- [7] Negri, M., Gamba, P., Lisini, G., Tupin, F.: Junction-Aware Extraction and Regularization of Urban Road Networks in High-Resolution SAR Images. *IEEE Transactions on Geoscience and Remote Sensing* 44(10) (2006)
- [8] Peteri, R., Celle, J., Ranchin, T.: Detection and Extraction of Road Networks from High Resolution Satellite Images. In: *Proceedings of the IEEE International Conference on Image Processing, Spain* (2003)
- [9] Shi, W., Zhu, C.: The Line Segment Match Method for Extracting Road Network From High-Resolution Satellite Images. *IEEE Transactions on Geoscience and Remote Sensing* 40(2) (2002)
- [10] Tupin, F., Maitre, H., Mangin, J.F., Nicolas, J.M., Pechersky, E.: Detection of Linear Features in SAR Images: Application to Road Network Extraction. *IEEE Transactions on Geoscience and Remote Sensing* 36(2) (1998)
- [11] Zhou, J., Bischof, W.F., Caelli, T.: Robust and efficient road tracking in aerial images. In: *CMRT 2005, IAPRS, Austria*, vol. 36 (2005)

Spatio-Temporal Filters in Video Stream Processing

Marek Szczepanski

Abstract. An overview of real-time video filtering techniques was presented in this paper. The extension of standard techniques based on temporal Gaussian combined with Fast Digital Paths Approach [8] was presented. Presented technique provides excellent noise suppression ability especially for low light sequences with low computational complexity.

1 Introduction

The widespread use of webcams, camcoders, digital cameras embedded in mobile phones allows capturing images and videos in different situations especially in low light environment.

Unfortunately, the high level of miniaturization of sensors entails very poor quality of recorded material. In addition, increasing the number of megapixels packed into such a small area in order to improve image quality imposes strong noise artifacts in resulting images. Similar problems with low light acquisition are encountered in medical imagery systems.

It is therefore necessary to use algorithms that improve the quality of such images. One of the most challenging tasks is shot noise removal which is dominant in low light images [3].

The speed of these processes is a critical parameter here. Their execution time should be less than the interval between consecutive frames.

In this paper, a novel noise spatio-temporal filter, which combines Temporal Gaussian smoothing with spatial Fast Digital Path Approach (*FDPA*) [8] has been proposed. Temporal Gaussian smoothing effectively removes shot noise or Gaussian artifacts but introduces ghosting artifacts in dynamic sequences, thus in such regions spatial filtering is used.

The paper is organized as follows. In Section 2 reference video filtering techniques were presented including spatial, temporal and spatio-temporal filters.

Marek Szczepanski

Institute of Automatic Control, Silesian University of Technology, Poland

e-mail: Marek.Szczepanski@polsl.pl

Section 3 introduces the concept of new filtering scheme, while Section 4 presents simulation results. Finally, Section 5 summarizes our paper.

2 Reference Video Filters

2.1 Spatial Filters

Spatial filtering uses standard image denoising techniques applied to subsequent frames. Spatial approach may be particularly effective for impulsive noise filtering, but as it does not use inter-frame information is not able to effectively remove Gaussian or shot noise. In this work following spatial filters were implemented:

- Linear convolution
- median filter
- Vector Median Filter (VMF)
- Fast Digital Paths Approach (FDPA) [8]

2.2 Fast Digital Paths Approach Spatial Filter

The FDPA Filter exploits possible connections between successive image pixels using the concept of digital paths in spatial domain. According to the proposed here methodology, image pixels are grouped together forming paths that reveal the underlying structural dynamics of the image.

In this work general fuzzy filtering structure proposed in [7] will be used. The general form of the fuzzy adaptive filters proposed in this work is defined as weighted average of input vectors inside the processing window W .

$$\hat{\mathbf{F}}_0 = \sum_{i=0}^{k-1} w_i \mathbf{F}_i = \frac{\sum_{i=0}^{k-1} \mu_i \mathbf{F}_i}{\sum_{i=0}^{k-1} \mu_i}, \quad (1)$$

where \mathbf{F}_i and $\hat{\mathbf{F}}_0$ denotes filter inputs and output respectively, μ is the similarity function calculated over digital paths of length n included in the processing window W .

The similarity function μ will be based on connection cost between two image pixels (x_0, y_0) and (x_n, y_n) connected by a digital path $P_m^W \{(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)\}$ of length n :

$$\chi_m^{W,n} \{(x_0, y_0), (x_n, y_n)\} = \sum_{k=0}^{n-1} \|\mathbf{F}(x_{k+1}, y_{k+1}) - \mathbf{F}(x_k, y_k)\|, \quad (2)$$

which plays the role of a measure of dissimilarity between pixels (x_0, y_0) and (x_n, y_n) , along a specific digital path P_m^W joining (x_0, y_0) and (x_n, y_n) , where m is the path index and $\|\cdot\|$ denotes the vector norm [9].

Let us now define a fuzzy similarity function between two pixels connected along all digital paths leading from (i, j) to (k, l) .

$$\mu^{W,n}\{(i, j), (k, l)\} = \sum_{m=1}^{\omega} \exp[-\beta \cdot \chi_m^{W,n}\{(i, j), (k, l)\}] \tag{3}$$

where ω is the number of all paths connecting (i, j) and (k, l) , β is a design parameter and $\chi_m^{W,n}\{(i, j), (k, l)\}$ is a total distance function along a specific path from a set of all ω possible paths joining (i, j) and (k, l) . In this way $\mu^{W,n}\{(i, j), (k, l)\}$ is a value, calculated over all routes linking the starting point (i, j) and the endpoint (k, l) .

In order to reduce filter complexity the FDPA filter uses fixed size of the supporting window W is set to (3×3) independently of the path's length.

2.3 Temporal Filters

Noise introduced by CCD and CMOS sensors significantly reduce the quality of the recorded material and cause considerable losses during compression. Because we are dealing with a sequence of images, rather than a single frame, we can eliminate the noise using inter frame relations.

It is important that the sensor noise is characterized by a low correlation between individual frames (with exception of hot pixels), while parts of the image, even the fast-changing, they are strongly correlated.

These properties are used during temporal filtering using different variants of averaging the values of individual pixels in successive video frames. The simplest temporal filter is the Temporal Arithmetic Mean Filter (*TAMF*), the output of that filter can be represented by the following relationships:

$$\hat{F}(i, j, t) = \frac{1}{n} \sum_{\Delta t=0}^{n-1} F(i, j, t + \Delta t), \text{ or} \tag{4}$$

$$\hat{F}(i, j, t) = \frac{1}{2n+1} \sum_{\Delta t=-n}^n F(i, j, t + \Delta t), \tag{5}$$

where F, \hat{F} denote the input and output frames and n determines temporal window size.

This method, although the simplest and quickest is only suitable for static sequences, because averaging frames where there are objects in motion creates "ghosting" effects in output sequence.

Figure 1 shows highly efficient filtering of static regions of sequence, with strong ghosting effect around moving objects. One way to reduce the ghosting effect is to use temporal Gaussian filtering instead of simply averaging:

$$\hat{F}(i, j, t) = \sum_{t=-n}^{t=n} g(\sigma, t) * F(i, j, t). \tag{6}$$



Fig. 1 Frame from standard video sequence *hall monitor* with additional Gaussian noise added ($\sigma = 10$) and the result of temporal averaging with window length $n = 9$.

Other simple approach involves movement detection so the filtering results can be described as follows:

$$\hat{F}(i, j, t) = \begin{cases} \frac{1}{n} \sum_{\Delta t=0}^{n-1} F(i, j, t + \Delta t), & \text{if } \sum_{\Delta t=0}^{n-1} |F(i, j, t + \Delta t) - F(i, j, t + \Delta t - 1)| \leq \epsilon \\ F(i, j, t) & \text{otherwise.} \end{cases} \quad (7)$$

Some algorithms utilize motion compensation to reduce blurring effect such as works presented by Dubois and Sabri [2]. Another solution to reduce ghosting artifacts is to use a temporal version of bilateral filter used as a element of ASTA filter [1].

2.4 Spatio-Temporal Filters

Spatio-temporal filtering is a combination of the methods discussed in previous chapters. Filter output is calculated using spatial neighbors of the processed pixel as well as its equivalents in other frames.

Using this information we can effectively improve the quality of the film, reducing negative filtration effects.

Larger set of input information needed to process one pixel means greater computational effort, among other reasons discussed in this chapter, the filters are among the most computationally complex and often they are used for off-line processing.

Established a whole group of filters which are a generalization of standard filters on the spatial-time-spatial domain. The principal advantage of this approach is its simplicity, in general no motion detection is needed.

These methods can include three-dimensional median filters are described in [5, 11]. Other solutions used for color sequences is three-dimensional version of the vector median filter (VMF3D).

Unfortunately, these methods often cause the formation of jagged edges of moving objects (Fig. 2).



Fig. 2 Frame from popular *Juggling* video sequence and the result of temporal three dimensional Vector Median Filtering (processing window: $3 \times 3 \times 3$, norm L_1).

This problem occurs to a lesser extent, by using an expanded time-space version of anisotropic diffusion of Perona and Malik [6] described by Suk Ho Lee and Moon Gi Kang [4].

Other very effective solution has been used in the ASTA filter (Adaptive Spatio-Temporal Accumulation Filter) [1], using a combination of temporal and spatial bilateral filter [10]. The ASTA filter and spatio-temporal anisotropic diffusion are very effective in noise reduction, however, require considerable computational effort.

3 Temporal Gaussian with Spatial FDPA Filter (*TGaussFDPA*)

As can be seen in Fig. 1 in the case of static scenes get excellent results using a simple averaging or Gaussian smoothing over time, which, however, completely does not work when the scene contains moving objects, then we should use the spatial filtering algorithms.

$$\hat{\mathbf{F}}(i, j, t) = \begin{cases} \sum_{t=-n}^{t=n} g(\sigma, t) * F(i, j, t), & \text{if} \\ \frac{1}{2n+1} \sum_{\Delta t=0}^{2n} \text{dist}(\mathbf{F}(i, j, t + \Delta t), \mathbf{F}(i, j, t + \Delta t - 1)) \leq \varepsilon \\ \mathbf{F}_{FDPA}(i, j, t), & \text{otherwise.} \end{cases} \quad (8)$$

where the distance between two points is calculated as module of the difference in luminance:

$$\text{dist}(\mathbf{F}_1, \mathbf{F}_2) = \left| \frac{1}{3} [(F_{1R} + F_{1G} + F_{1B}) - (F_{2R} + F_{2G} + F_{2B})] \right| \quad (9)$$

Figure 3 b) presents temporal Gaussian filtering with movement detection mask calculated from the formula 9. The selected pixels were completely ignored by the filter, since the differences between their counterparts in the subsequent frames were too big. This example shows a flaw presented approach, a mask containing pixels skipped in the processing time is often torn and heterogeneous. This can lead to noticeable artifacts in the vicinity of moving objects.

The solution to this problem would be to carry out the opening of our mask. The results of this operation are shown in 3 c) and d).



Fig. 3 a) Frame from standard video sequence *hall monitor* with additional Gaussian noise added ($\sigma = 10$), b) Temporal Gaussian filtering with mask of skipped pixels, c) mask of skipped pixels after morphological opening (with mask of size 3×3) and d) Final filtering result ($n = 5, \sigma = 5, \varepsilon = 11, \beta = 15$).

4 Simulation Results

Several filters capable of real-time video processing were examined on numerous video sequences. Subjective results were obtained from original noisy video sequences, however some synthetic tests with artificial noise were also performed.

Objective quality measures such as the *Root Mean Squared Error* (RMSE), the *Signal to Noise Ratio* (SNR), the *Peak Signal to Noise Ratio* (PSNR), the *Normalized Mean Square Error* (NMSE) and the *Normalized Color Difference* (NCD) were used for the analysis. All those objective quality measures were calculated for the sequence of the filtered images.

The performance of the following filters was evaluated:

- Temporal Gaussian Filter *TGauss* (with time window $n = 5$ and $\sigma = 5$),
- Spatial Vector Median Filter (with window 3×3 and L_1 norm),
- Spatial Fast Digital Paths Approach *FDPA* ($\beta = 15$),
- Spatio-temporal Vector Median Filter - *VMF3D* (with window $3 \times 3 \times 3$ and L_1 norm),
- Temporal Gaussian Filter combined with Fast Digital Paths Approach *TGaussFDPA* ($n = 5, \sigma = 5, \varepsilon = 11, \beta = 15$)

Figure 4 shows the frame from noisy sequence captured in low light conditions containing small toy-car moving rapidly. It can be seen that temporal methods produces perfect background while moving object is blurred, static techniques can't clear the noise effectively. Only combination of spatial and temporal techniques gives satisfactory results.

Objective quality measures for sequence *Foreman* corrupted with Gaussian noise ($\sigma = 20$) are presented in Table 1 while Table 2 combines results of filtering *Hall monitor* sequence with Gaussian noise ($\sigma = 10$). Average values and their standard deviation are presented their standard deviation were evaluated.

Table 1 Comparison of the filtering algorithms applied for *Foreman* sequence corrupted with Gaussian noise ($\sigma = 20$)

Filter	PSNR [dB]	σ_{PSNR} [dB]	NCD [10^{-4}]	σ_{NCD} [10^{-4}]
None	22.22	0.03	185.8	24.7
VMF	25.7	0.72	109.4	16.5
VMF3D	26.6	1.81	80.0	11.6
FDPA	30.3	1.75	47.7	7.3
TGauss	25.7	2.23	90.2	11.9
TGaussFDPA	28.2	0.51	80.1	11.3

It can be noticed that the sequence *Hall monitor* has mean PSNR value slightly better for temporal Gaussian filter then for the hybrid one (2). However for temporal

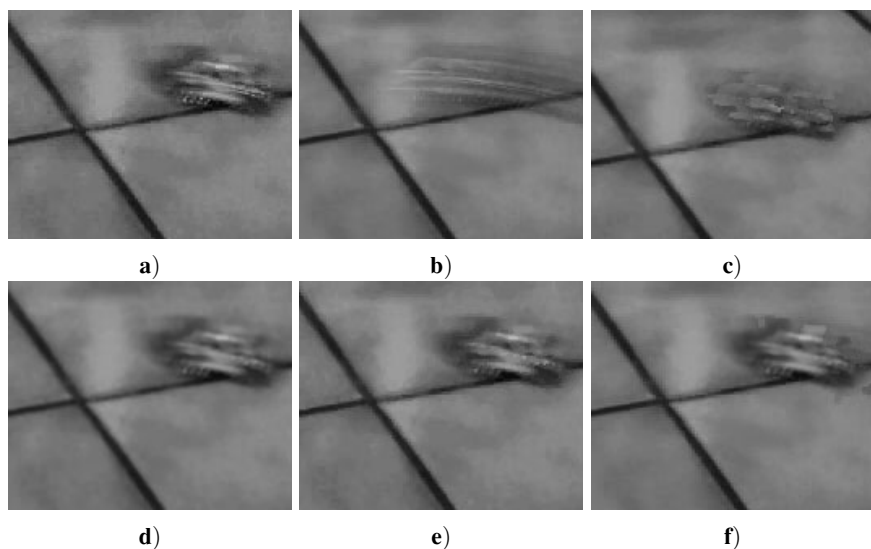


Fig. 4 a) Frame from the noisy video sequence - *Toy Car* and results of filtering with b) Temporal Gaussian, c) *VMF3D*, d) *Spatial FDPA*, e) *VMF* and f) Temporal Gaussian with *Spatial FDPA*.

Table 2 Comparison of the filtering algorithms applied for *Hall monitor* sequence corrupted with Gaussian noise ($\sigma = 10$)

Filter	PSNR [dB]	σ_{PSNR} [dB]	NCD [10^{-4}]	σ_{NCD} [10^{-4}]
None	28.3	0.01	95.2	10.4
VMF	28.8	0.57	62.2	6.9
VMF3D	29.4	0.14	51.0	11.1
FDPA	32.4	0.05	38.4	4.3
TGauss	33.5	0.79	51.2	13.2
TGaussFDPA	32.9	0.09	50.4	8.4

filtering gives much higher standard deviation. The Figure 5 depicts PSNR values of subsequent frames filtered with temporal Gaussian and spatio-temporal filter, it can be clearly seen when temporal filters produces ghosting artifacts.

One of the key issues in the processing video sequence is real-time computational efficiency. The average speed of various algorithms for test sequences *Hall monitor* and *Foreman* with a CIF resolution (352x288) is presented in Table 3. Tests were conducted on a computer with an Intel Core i7 running at 2.67 GHz with single processing thread. All filters based on temporal Gaussian use the time window length $n = 5$ and $\sigma = 5$, the filters with motion detection uses constant threshold $\epsilon = 11$.

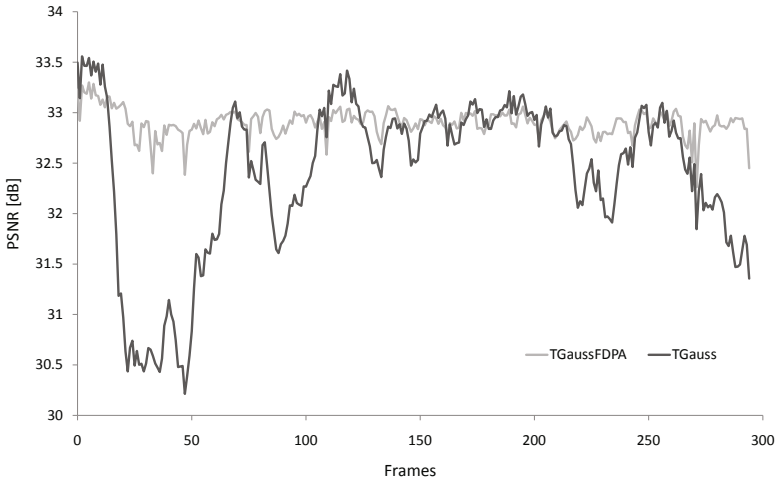


Fig. 5 PSNR coefficients of noisy *Hall monitor* sequence with TGauss and TGaussFDPA filters

The fastest is of course the simplest one - temporal Gaussian, while the slowest FDPA and VMF3D is rather complicated for real-time processing. This results can be improved by utilizing additional threads, however the gain is not linear.

Table 3 The average processing speeds for test sequences *Hall monitor* and *Foreman*

Filter	<i>Hall monitor</i> [fps]	<i>Foreman</i> [fps]
VMF	27.2	26.1
VMF3D	4.7	4.7
FDPA	3.5	3.5
TGauss	71	71
TGaussFDPA	37.5	15.8

5 Conclusions

From several years we can observe increasing interest in video processing. Video noise reduction without structure degradation is perhaps the most challenging video enhancements task. Several techniques have been proposed over the years. Among them are standard noise reduction techniques, the so-called spatial filters, applied to subsequent frames of the video stream. However, standard image processing techniques cannot utilize all available information i.e. similarities in neighboring frames, so modern video denoising algorithms utilize also temporal information.

An overview of real-time video filtering techniques was presented in this paper. The new approach is based on temporal Gaussian combined with movement detection and spatial filtering was presented. Presented technique provides excellent noise suppression with low computational complexity.


Acknowledgements. This work has been supported by the Polish Ministry of Science and Higher Education under R&D grant no. N N516 374736 from the Science Budget 2009-2011.

References

- [1] Bennett, E.P., McMillan, L.: Video enhancement using per-pixel virtual exposures. *ACM Trans. Graph.* 24(3), 845–852 (2005)
- [2] Dubois, E., Sabri, S.: Noise reduction in image sequences using motion-compensated temporal filtering. *IEEE Transactions on Communications* 32(7), 826–831 (1984)
- [3] Lee, S., Maik, V., Jang, J., Shin, J., Paik, J.: Noise-adaptive spatio-temporal filter for real-time noise removal in low light level images. *IEEE Transactions on Consumer Electronics* 51, 648–653 (2005)
- [4] Lee, S., Kang, M. G.: Spatio-temporal video filtering algorithm based on 3-d anisotropic diffusion equation. In: *Proceedings of International Conference on Image Processing, ICIP 1998*, vol. 2, pp. 447–450 (1998)
- [5] Meguro, M., Taguchi, A., Hamada, N.: Data-dependent weighted median filtering with robust motion information for image sequence restoration. In: *Proceedings of International Conference on Image Processing, ICIP 1999*, vol. 2, pp. 424–428 (1999)
- [6] Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 629–639 (1990)
- [7] Plataniotis, K., Androustos, D., Venetsanopoulos, A.: Fuzzy adaptive filters for multi-channel image processing. *Signal Processing Journal* 55(1), 93–106 (1996)
- [8] Szczepanski, M., Smolka, B., Plataniotis, K., Venetsanopoulos, A.: On the geodesic paths approach to color image filtering. *Signal Processing* 83(6), 1309–1342 (2003)
- [9] Toivanen, P.: New geodesic distance transforms for gray scale images. *Pattern Recognition Letters* 17, 437–450 (1996)
- [10] Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: *ICCV*, pp. 839–846 (1998)
- [11] Viero, T., Neuvo, Y.: Non-moving regions preserving median filters for image sequence filtering. In: *IEEE International Conference on Systems Engineering*, pp. 245–248 (1991)

Simplifying SURF Feature Descriptor to Achieve Real-Time Performance

Marek Kraft and Adam Schmidt

Abstract. The detection and matching of interest points serves as the base for many computer vision algorithms, such as visual odometry, structure from motion, tracking or simultaneous localization and mapping. The accuracy of matching is therefore of very high importance. This requirement is however often irreconcilable with the requirement of real-time performance, especially on resource constrained architectures. In this paper, we analyze a few possible simplifications to the recently developed SURF feature description and matching scheme, enabling it to shorten the processing time on virtually all computing platforms. The introduced simplifications do not introduce any significant matching performance penalty when compared with the full SURF implementation in the aforementioned applications .

1 Introduction

Numerous modern computer vision applications require robust feature detection and matching algorithms. Over the years, the methods based on detection of natural features have superseded marker based methods. The methods based on natural features require the detection of distinctive image keypoints and matching them across different scene views. Active research in the field of feature detection and description led to the development of numerous algorithms, that can successfully cope with realistic application scenarios. One of the most recently developed algorithms is the SURF multiscale feature detector and descriptor.

In this paper, we analyze the possible simplifications, that can be introduced to the SURF feature descriptor, to achieve real-time performance of the algorithm.

Marek Kraft · Adam Schmidt

Poznań University of Technology, Institute of Control and Information Engineering,
Piotrowo 3A, 60-965 Poznań, Poland

e-mail: Marek.Kraft@put.poznan.pl

¹ This project is partially funded by the Polish Ministry of Science and Higher Education, project number N N514 213238.

The research presented in [13] has shown, that for indoor mobile robot navigation applications, the high-performance SURF feature descriptors can be paired with fast, single-scale feature detectors like Harris [8] or FAST [12] with very good results. The simplified version of SURF is faster, enabling real-time performance on resource constrained architectures, with just a minor performance penalty. Furthermore, it can also be easily adopted to dedicated hardware architectures.

2 SURF Feature Descriptor

The recently developed SURF feature descriptor [1] is built upon the SIFT feature descriptor [11]. Both of these multiscale descriptors encode the distribution of pixel intensities in the neighborhood of the detected feature, rather than considering raw pixel intensities as it is the case with commonly used similarity measures like sum of absolute differences (SAD), sum of squared differences (SSD), normalized cross-correlation (NCC) etc. In the case of SIFT, the distribution of pixel intensities is encoded as gradient direction histograms, while the SURF descriptor uses gradient responses computed in the x and y direction using Haar wavelets – see figure 1.



Fig. 1 Haar wavelets used to compute the gradient in the x and y directions, respectively. The weights are equal to 1 for the white areas, and -1 for the black areas.

The use of Haar wavelets in conjunction with integral images allows to decrease computation time when compared to SIFT [10]. Computation time does not depend on the size of the wavelet, so the scale space is built by resizing the feature instead of resizing the original image, resulting in further speedup. Computation of the descriptor for a given feature can be divided into two distinct tasks.

The first task is the dominant orientation assignment, necessary to achieve rotation invariance. To this goal, the Haar wavelet responses in x and y directions are computed for all points located within the circle with the radius of $6s$, centered at the interest point, where s is the scale at which the interest point was detected. The wavelets are also properly resized to correspond with the scale s – the length of the side is set to $4s$. The responses are then weighted with a Gaussian centered at the interest point. The standard deviation of the Gaussian is also scale dependent – $\sigma = 2s$. The response at each one of the points gives rise to a point in the vector space, with the magnitude of the x -response being the value on the abscissa, and the magnitude of the y -response being the value on the ordinate. The dominant orientation is selected by rotating a circle sector covering an angle of $\frac{\pi}{3}rad$ around the interest point. The responses (the x and y gradient values) in each segment are summed and form a resultant vector. The angle of the longest resultant vector is assigned as the dominant orientation for a given interest point.

The computation of the descriptor itself begins with placing a square window with the size length of $20s$, so that its center of this window is aligned with the interest point, and its orientation is taken from the previous step. This window is subsequently divided into 4×4 square subregions. Inside these subregions, 5×5 regularly spaced sample points are selected. Consult figure 2 for an illustration.

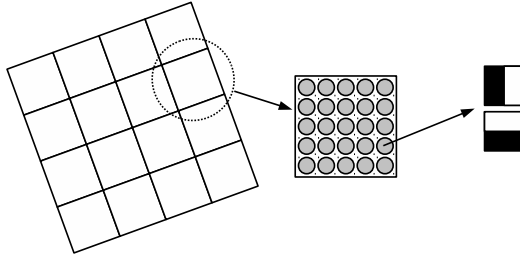


Fig. 2 Structure of the window used for the computation of the SURF descriptor for a given feature (as used in SURF64).

In the following step, Haar wavelet responses are computed at each one of these sample points. The responses (denoted dx and dy) are computed along the two principal directions using the masks show in figure 1, with the mask side length of $2s$, and then weighted with a Gaussian, centered at the center of the subregion ($\sigma = 2.5s$) to increase the robustness to geometric deformations and localization errors. The complete form of the descriptor for each of the subregions is computed by summing the responses and their absolute values acquired at each of the sample points (see equation 1). The sums are additionally weighted with a 2D Gaussian according to the subregion coordinates ($\sigma = 1.5$).

$$DESC_{sub} = [\sum dx, \sum dy, \sum |dx|, \sum |dy|] \quad (1)$$

Incorporating the sums of absolute values of the responses into the descriptor allows for encoding of more complex intensity patterns, making the descriptor more distinctive. Every subregion adds 4 elements to the descriptor. Hence, for 16 subregions, the descriptor size is 64 elements, and the descriptor name is SURF64. To achieve the invariance to contrast, the summed values are normalized to a unit length vector. The authors proposed also, among other others, the reduced version of the descriptor called SURF36. In this variant, the window processed is divided into 3×3 square subregions, resulting in reduced dimensionality. This results in shorter computation time and faster matching. If the rotational invariance is not required, the upright version of the descriptor, called USURF can be used. Skipping the dominant orientation assignment step yields further reduction of the time required for computation. The authors report, that such simplified version is robust to in-plane rotations in the range of $\pm 15^\circ$.

3 Changes Introduced to the Original Algorithm

Three different modifications were introduced to the SURF descriptor in order to reduce its complexity and computational requirements. One of our goals was to check whether the Gaussian weighting of the sample points' responses and subregions' descriptors has noticeable influence on the matching reliability. Therefore, disabling the Gaussian weighting was the first proposed modification of the SURF descriptor.

The elements of the original SURF descriptor are stored as an array of 32-bit floating point numbers. As the descriptors are compared by the means of the Euclidean distance switching to the integer representation can significantly reduce the time and resources needed for both the computation and comparison of the distance between two descriptors. Therefore, the components of the modified descriptor were stored as an array of integers.

Such change required reformulation of the descriptor normalization. The original SURF descriptor is scaled to a unit length vector which means that the values of components lie in the range of $\langle -1.0, 1.0 \rangle$. In case of the integer representation the minimum and maximum value of descriptor's components depend of the number of bits used to represent the integer ($SURF_i \in \langle -2^{n-1}, 2^{n-1} - 1 \rangle$, where $SURF_i$ is the i -th component and n is the number of bits). Instead of normalizing to a unit scale vector the integer descriptors were scaled to have a length equal to 2^{n-1} .

The last change introduced to the SURF descriptor was aimed at reducing the complexity of the descriptors normalization. Scaling to a unit length vector requires calculating the length of the 36 or 64 dimensional vector (involving calculation of the square root). The normalization was considerably simplified by ensuring that the maximal absolute value of the descriptor's components is equal to 2^{n-1} .

4 Experiment Description

The unaltered, floating point versions of the USURF36 and USURF64 served as the base for the simplifications evaluation [6]. The first part of the experiment was aimed at evaluating the influence of Gaussian weighting on the matching performance. Therefore, the variants of the floating point USURF36 and USURF64 descriptors with the disabled weighting of sample points, subregions or both of them were used. After that the performance of the integer USURF36 and USURF64 descriptors for different sizes of integers ($n \in \langle 8, 10, 12, 14, 16 \rangle$) and both of the above mentioned normalization methods was evaluated. No Gaussian weighting was used with the integer descriptors.

All 28 variants of the SURF descriptor were evaluated on three different video sequences. The first sequence was registered in the lab room and is distinguished by high feature density. The second sequence should allow to test the matching performance on poorly textured image sequence with low feature density and was recorded in a corridor with uniformly colored walls and floor. The third sequence is an urban outdoor sequence, recorded in a car park area. Due to a fairly rough, gravel-like ground structure, the sequence displays significant camera shake during movement.

The UI-1225LE-C uEye camera (IDS Imaging) with 1/3" CMOS sensor was used for registration. The camera enables fast registration (up to 87 FPS) with the global shutter. The camera was equipped with a wide-angle lens ($f = 2.8mm$). The images were undistorted according to camera model presented in [4]. The camera parameters were obtained using calibration functions available in the OpenCV library [3]. To maintain stable framerate, exposure time was set to a constant value, but the automatic gain control function was used. No filtering of images was performed, therefore the sequences display many realistic effects – noise, contrast variations etc.



Fig. 3 Exemplary images from the registered sequences

To evaluate the influence of algorithm simplifications on matching performance, each of the earlier described variants of the descriptor was tested on all three sequences. For each sequence the descriptors performance was evaluated while matching consecutive frames, every fifth and every tenth frame which simulated lower camera frame rate. No restrictions were made to the search area and no cross checking was performed. The features pair was considered to be matched if the distance between their descriptors was less than 0.65 of the distance between the second-best pair.

To evaluate the quality of matches, we used the 8-point algorithm for fundamental matrix computation [9], with robust estimation method based on the RANSAC (random sample consensus) algorithm [7], readily available as a function in the OpenCV computer vision software library [3]. RANSAC is an iterative method allowing for the estimation of a mathematical model (the fundamental matrix in our case) from a set of observations (matched points) containing outliers (false matches). We chose the ratio of the number of inliers (i.e. feature matches consistent with the estimated



Fig. 4 Exemplary frame before and after image rectification

Table 1 The average inliers to matched features ratio of the standard USURF descriptors

Descriptor length	64	36
Seq. 1 step 1	0.88	0.88
Seq. 1 step 5	0.82	0.80
Seq. 1 step 10	0.78	0.75
Seq. 2 step 1	0.83	0.83
Seq. 2 step 5	0.74	0.74
Seq. 2 step 10	0.78	0.75
Seq. 3 step 1	0.93	0.94
Seq. 3 step 5	0.89	0.90
Seq. 3 step 10	0.87	0.88

fundamental matrix) to the number of all the detected matches as the metric allowing for the evaluation of the quality of the matching process. Any case in which the number of inliers was less than 12 was treated as a matching failure, so that at least 4 additional point pairs (aside from the original 8 pairs used to compute the fundamental matrix) must be consistent with the generated model. The symmetric reprojection error [9] was used to check consistency of the data with mathematical model.

5 Results

Table 1 presents the ratio of inliers to the total number of matched features for the standard USURF36 and USURF64 descriptors. The length of the descriptor had no significant impact on the matching reliability. Descriptors performed worse while matching every fifth and every tenth frame which was caused by bigger differences between images.

Generally, the Gaussian weighting of the sample points and subregion components did not have notable influence of the average inliers to matched features ratio (see Table 2). The performance dropped only during matching of the every tenth frame of the first two sequences. However, the performance drop could be compensated by weighting only either sample points or subregions.

Table 2 The inliers to number of matched features ratio of floating point USURF descriptor

Descriptor length	64	64	64	36	36	36
Sample weighting	+	-	-	+	-	-
Subregion weighting	-	+	-	-	+	-
Seq. 1 step 1	0.88	0.88	0.88	0.88	0.88	0.88
Seq. 1 step 5	0.81	0.81	0.81	0.81	0.80	0.80
Seq. 1 step 10	0.77	0.75	0.76	0.75	0.75	0.72
Seq. 2 step 1	0.83	0.82	0.82	0.83	0.83	0.83
Seq. 2 step 5	0.73	0.75	0.74	0.72	0.73	0.73
Seq. 2 step 10	0.77	0.75	0.76	0.75	0.75	0.72
Seq. 3 step 1	0.92	0.93	0.92	0.94	0.94	0.93
Seq. 3 step 5	0.89	0.89	0.88	0.90	0.90	0.89
Seq. 3 step 10	0.86	0.86	0.86	0.87	0.87	0.87

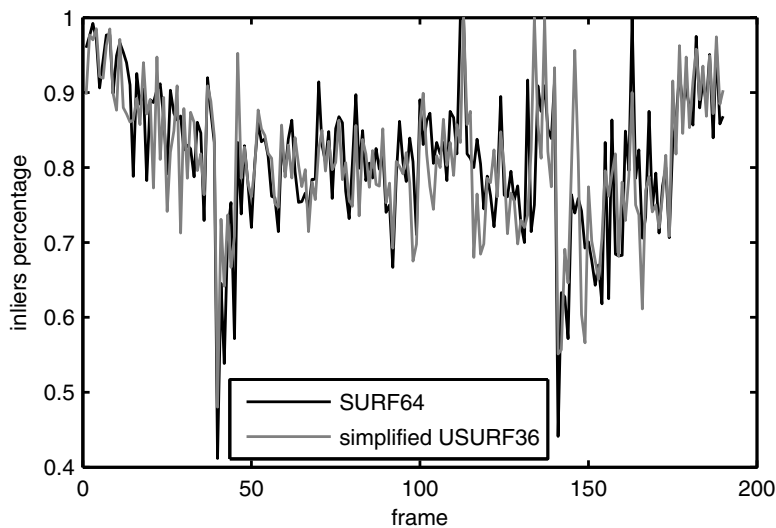
Table 3 The inliers to number of matched features ratio of integer USURF descriptor with unit length scaling

Descriptor length	64	64	64	64	64	36	36	36	36	36
Number of bits	16	14	12	10	8	16	14	12	10	8
Seq. 1 step 1	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
Seq. 1 step 5	0.81	0.81	0.81	0.81	0.81	0.80	0.80	0.80	0.80	0.80
Seq. 1 step 10	0.76	0.76	0.77	0.76	0.76	0.73	0.73	0.73	0.73	0.73
Seq. 2 step 1	0.82	0.82	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.83
Seq. 2 step 5	0.74	0.74	0.74	0.74	0.73	0.73	0.73	0.73	0.73	0.73
Seq. 2 step 10	0.76	0.76	0.77	0.76	0.76	0.73	0.73	0.73	0.73	0.73
Seq. 3 step 1	0.92	0.92	0.92	0.92	0.92	0.93	0.93	0.93	0.93	0.93
Seq. 3 step 5	0.88	0.88	0.88	0.88	0.88	0.89	0.89	0.89	0.89	0.89
Seq. 3 step 10	0.86	0.86	0.86	0.86	0.86	0.87	0.87	0.87	0.87	0.87

Tables 3 and 4 present the performance of the integer-based USURF36 and USURF64 descriptors without the Gaussian weighting of the sample points and the subregions. In case of the results in Table 3, the descriptors were normalized as unit length vectors. Table 4 shows the performance of the descriptors normalized according to the maximal component. No significant difference between the performance of the floating-point and the integer descriptors was observed. The simplification of the descriptor normalization not only did not reduce the matching reliability, but actually slightly improved it.

Table 4 The inliers to number of matched features ratio of integer USURF descriptor with maximal scaling

Descriptor length	64	64	64	64	64	36	36	36	36	36
Number of bits	16	14	12	10	8	16	14	12	10	8
Seq. 1 step 1	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
Seq. 1 step 5	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81
Seq. 1 step 10	0.77	0.77	0.77	0.77	0.78	0.75	0.75	0.75	0.75	0.75
Seq. 2 step 1	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
Seq. 2 step 5	0.74	0.74	0.74	0.74	0.74	0.73	0.73	0.73	0.73	0.73
Seq. 2 step 10	0.77	0.77	0.77	0.77	0.78	0.75	0.75	0.75	0.75	0.75
Seq. 3 step 1	0.92	0.92	0.92	0.92	0.92	0.94	0.94	0.94	0.94	0.94
Seq. 3 step 5	0.88	0.88	0.88	0.89	0.88	0.90	0.90	0.90	0.90	0.90
Seq. 3 step 10	0.86	0.86	0.86	0.86	0.86	0.88	0.87	0.87	0.87	0.88

**Fig. 5** An example time course of the percentage of inliers for SURF64 and simplified USURF36 over a whole sequence.

An example time course of the percentage of inliers for SURF64 and simplified USURF36 over a whole sequence (sequence number 1 with frame skip of 5 in this case) further proves, that the differences between the full SURF and its truncated version are minimal in the considered benchmarks.

6 Conclusions

This paper presented a number of modifications of the SURF descriptor aimed at reducing the numerical complexity and computational requirements. It was experimentally shown that the Gaussian weighting of sample point responses and sub-region descriptors can be simplified. Moreover, it is possible to switch to the 8-bit integer representation of the descriptor components and to simplify the normalization process without any effect on the feature matching reliability. The alterations to the original SURF descriptors considerably reduce the computational power necessary to match features between images. This is especially important in applications running on resource-constrained platforms which, as opposed to complex desktop PCs, sacrifice computing power for lower total system cost, power consumption and complexity. Such applications include mobile computing (smartphones, tablets), mobile robotics, automotive applications etc. Switching from floating point to integer arithmetics allows to reduce the complexity and resource cost of implementation of the algorithm on custom computing machines like FPGAs (field programmable gate arrays), ASICs (application-specific integrated circuits) and embedded systems [5]. Applications running on the common PCs can also benefit from these simplifications. Proper optimization of the code allows for the use of SIMD (single instruction, multiple data) extensions, implemented in modern x86-compatible microprocessors with so called 'packed instructions' [2]. These instructions allow for, among other things, the simultaneous, parallel operation on a set of data held inside 128-bit wide dedicated registers. One such register can therefore hold sixteen 8-bit integer values, but only four 32-bit floating point values. Packed integer arithmetic instructions are also faster than floating-point instructions. All this results in a significant reduction of processing time.

References

- [1] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). *Computer Vision and Image Understanding* 110(3), 346–359 (2008)
- [2] Bik, A.J.C., Kreitzer, D.L., Tian, X.: A Case Study on Compiler Optimizations for the Intel Core 2 Duo Processor. *International Journal of Parallel Programming* 36(6), 571–591 (2008)
- [3] Bradski, G., Kaehler, A.: *Learning OpenCV*. O'Reilly Media Inc., Sebastopol (2008)
- [4] Brown, D.: Close-range camera calibration. *Photogrammetric Engineering* 37(8), 855–866 (1971)
- [5] Deschamps, J.-P., Bioul, G.J.A., Sutter, G.D.: *Synthesis of Arithmetic Circuits: FPGA, ASIC and Embedded Systems*. Wiley-Interscience, Hoboken (2006)
- [6] Evans, C.: Notes on the OpenSURF library. Technical Report CSTR-09-001, University of Bristol (2009)
- [7] Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), 381–395 (1981)
- [8] Harris, C., Stephens, M.: A combined corner and edge detector. In: *Proceedings of the 4th Alvey Vision Conference*, pp. 147–151 (1988)

- [9] Hartley, R.I., Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press, Cambridge (2004) ISBN: 0521540518
- [10] Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. In: *Proc. of International Conference on Image Processing*, vol. (1), pp. 900–903 (2002)
- [11] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
- [12] Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
- [13] Schmidt, A., Kraft, M., Kasinski, A.J.: An evaluation of image feature detectors and descriptors for robot navigation. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) *ICCVG 2010*. LNCS, vol. 6375, pp. 251–259. Springer, Heidelberg (2010)

Algorithm for Accurate Determination of Contact Angles in Vision System for High-Temperature Measurements of Metals and Alloys Surface Properties

Tomasz Koszmider, Krzysztof Strzecha, Anna Fabijańska, and Marcin Bakala

Abstract. In this paper problem of contact angle determination using image processing and analysis algorithms is considered. Specifically, algorithms developed for computerised system for high temperature measurements of surface properties of metals and alloys are introduced. Application of the proposed methods allows to overcome limitations caused by specimen intense thermal radiation and obtain accurate values of surface tension in a wide range of temperatures.

1 Introduction

In many technological, high-temperature processes i.e. welding, composite material making, powder sintering, saturation of porous structures, coating etc. an important role is played by the phenomena occurring in contact of the liquid and the solid phase [1, 2, 3, 4, 5]. Therefore, investigation of physical-chemical processes occurring between a liquid and a solid is an important technological problem.

The main measurable quantities characterizing interactions between the solid and the liquid are surface tension of the liquid phase and the extreme angle of wetting of the base by a liquid [6, 7, 8]. Once the above quantities have been measured, it is possible to determine the remaining important quantities of the system, such as the adhesion energy, interfacial tension and, possibly, adsorption.

In spite of their importance, the surface properties of metals and alloys have not been fully investigated. This is because their measurements are performed in very high temperatures and are related to different problems arising from the activity of liquid metals and sensitivity of surface phenomena to impurities and the measurement conditions (temperature, pressure, atmosphere etc.). Therefore, recent results vary, depending on the method used for the measurements.

Tomasz Koszmider · Krzysztof Strzecha · Anna Fabijańska · Marcin Bakala
Computer Engineering Department,
Technical University of Lodz,

18/22 Stefanowskiego Str., 90-924 Lodz, Poland

e-mail: tkoszmi, strzecha, an_fab, mbakala@kis.p.lodz.pl

In order to provide reliable results and create a stable measurement environment the computerised system for high temperature measurements of surface properties was build in Computer Engineering Department of Technical University of Lodz (Poland) [9, 10]. In the system, the wetting angles and surface tension of metals and alloys are determined by image processing and analysis algorithms [11, 12].

The method commonly used for specimen shape description and surface properties determination is Axisymmetric Drop Shape Analysis (ADSA). The method however is not free from imperfections. The main weaknesses of the method are its computational complexity and difficulty of implementation. In this paper we propose the alternative solution which eliminates main drawbacks of ADSA. The algorithm was developed for the considered measurement system and significantly improves the quality of obtained results.

This paper is organised as follows. Firstly, in Section 2 the description of Thermo-Wet hardware and software is given. It is followed in Section 3 by description of the improved algorithm for contact angle determination. Results are given in Section 4. Finally, Section 5 concludes the paper.

2 The Experimental Setup

The paper presents the results of research on images acquired from computerized system for high temperature measurements of surface properties (wetting angle and surface tension) of liquid and solid in contact. The system named Thermo-Wet is capable of measuring the surface tension of a liquid and the wetting angle of a solid by a liquid over a wide range of temperatures (up to 1800°C).

2.1 Hardware

The apparatus of Thermo-Wet system consists of:

1. vision unit with a CCD camera, image analysis and processing algorithms and a set of infrared filters with the algorithm of their automatic changes;
2. specimen insertion mechanism;
3. technological gases supply system and system for the precise temperature measurement and control;
4. computer control of the measurement process;
5. high-temperature electric furnace with a protective atmosphere.

The block diagram of Thermo-Wet is presented in Figure 1.

2.2 Image Processing Algorithms

Crucial part of the measurements is connected with vision unit of Thermo-Wet. The unit is responsible for the following operations:

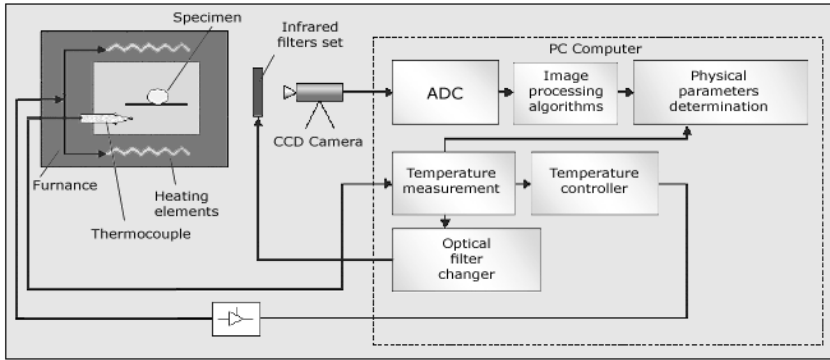


Fig. 1 Block diagram of Thermo-Wet system

1. image acquisition;
2. image preprocessing and segmentation;
3. image analysis, localisation of the specimen and determination of its geometrical parameters;
4. determination of specimen surface properties: contact angles and surface tension.

The considered vision system determines the surface properties of metals and alloys using the sessile drop method [6]. The method relates the characteristic dimensions of a drop of molten material into values of surface tension. Specifically, drop maximum width X and drop maximum height H are related to surface tension through the Equation (1).

$$\gamma = g \Delta \rho a^2 \tag{1}$$

where:

- γ - surface tension;
- g - gravity acceleration;
- $\Delta \rho$ - gradient of densities;
- a - parameter determined using Equation(2).

$$\frac{a^2}{X^2} = \left(\frac{H}{X}\right)^2 - 0.66 \left(\frac{H}{X}\right)^3 [1 - 4.05 \left(\frac{H}{X}\right)^2] \tag{2}$$

Drop characteristic dimensions are shown in Figure 2

In order to determine the left and the right contact angle (denoted as α_1 and α_2 respectively) it is necessary to:

1. approximate specimen shape;
2. define intersection points of approximated shape and the upper edge of the base plate;
3. compute angles between the base and tangents to specimen shape in the intersection points.

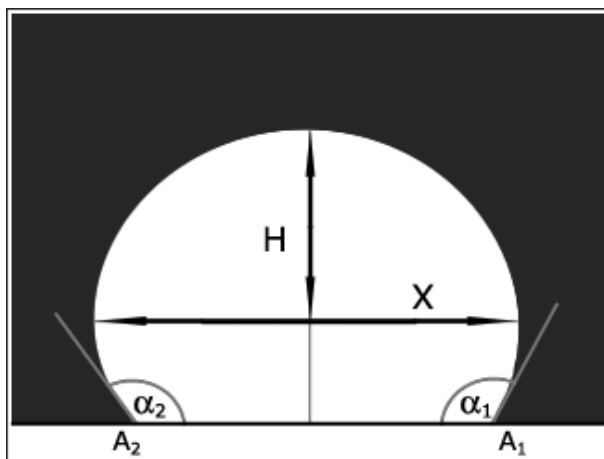


Fig. 2 Drop characteristic parameters.

A more detailed description of the algorithms recently used in Thermo-Wet system for contact angles determination can be found in [11].

In the considered system, the determination of contact angles is hindered by factors resulting from high temperature of the measurements and usage of gas protective atmosphere. Specifically, aura phenomenon and specimen reflection on the base should be mentioned. The method which allows to overcome the problems caused by these factors is introduced in the next section.

3 Contact Angle Determination Algorithm

Contact angle is the angle between the base plate and the tangent to the specimen profile in contact of three phases: the solid phase of base plate, the liquid phase of a specimen and the gas phase of the measurement atmosphere. In Thermo-Wet system the proper determination of contact angles is hindered by an intense illumination of the observed specimen. The illumination is a result of specimen thermal radiation (measurements are performed in temperatures of 1000°C - 1800°C) and causes the deformation of the specimen shape in contact of three phases. The deformation is not a result of wetting the base what was verified experimentally. The problem is shown in Figure 3. The specimen profile remains rounded while there should be the distinct angle between the specimen profile and the base plate.

In order to diminish the influence of the abovementioned deformation on the determined values of contact angles it is necessary to find all points located very close to the base plate but not belonging to the deformed part of the profile. Description of this algorithm can be found in [11]. The search is performed for both:

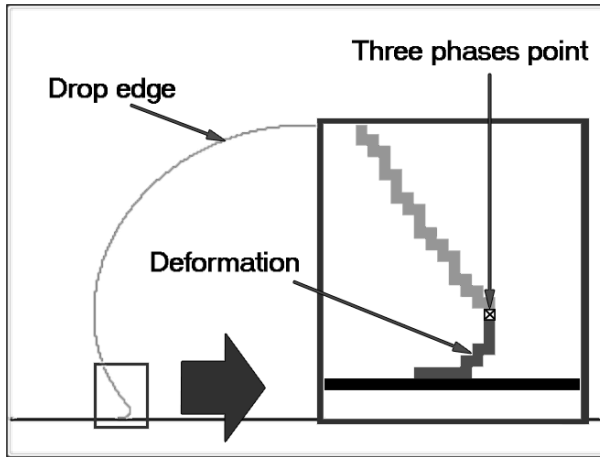


Fig. 3 The deformation of specimen shape profile in contact of three phases.

the left and the right side of the specimen profile. In order to determine the tangent to the specimen profile and the tangent angle to the base plate its necessary:

- to expand the set of edge points, so that the point of contact of three phases is not the utmost point of the approximated set;
- to approximate the extended set of points using polynomial of 5th or 6th order (depending on the shape of the drop);
- to determine the tangent to the curve approximating specimen profile in the point of contact of three phases.

The determined angle is equal to the contact angle.

The set of edge points of each profile after excluding points corresponding to the deformed part of the profile is approximated by a polynomial of 5th or 6th order. After the approximating curve is determined the tangent to the specimen profile is found in contact of three phases. The tangent becomes an abscissa axis OY' in local coordinate system with a center in point of contact of three phases. It is shown in Figure 4.

Extending the set of edge points is simply determining points A' axially symmetric to axis OX' of local coordinate system. Points are added to the set of points located above point of contact of three phases A .

Knowing the equations of lines along axes OX' and OY' in the local coordinate system and the coordinates of point A , it is possible to designate the coordinates x' and y' as the distance from the axis OX' and the axis OY' respectively. Determining coordinates of additional points in the global coordinate system is equivalent to solving the set of equations for local coordinates x' and y' .

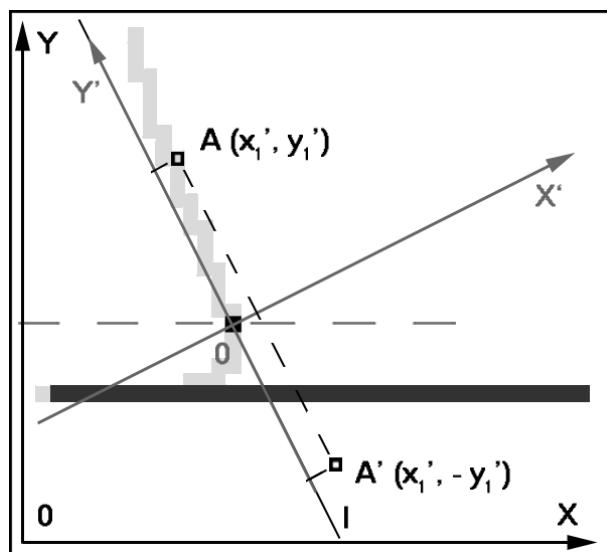


Fig. 4 The method of determining additional points of specimen shape profile.

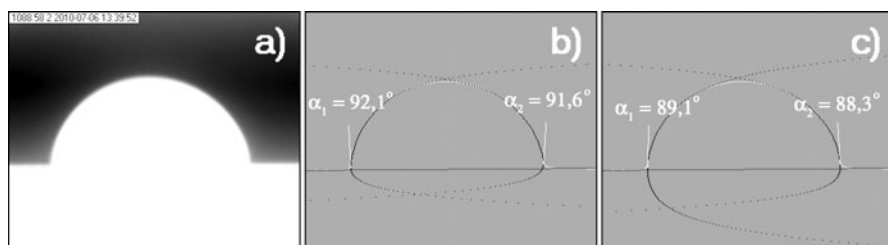


Fig. 5 Results of contact angle determination in exemplary image; (a) original image, steel, 1088 °C on Al_2O_3 surface; (b) results obtained using the original profile, (c) results obtained after extending the specimen profile using the proposed method.

4 Results

Results of contact angle determination in exemplary image of steel at temperature 1088°C are shown in Figure 5. Specifically, Figure 5a shows the original image. In Figure 5b results obtained using the original specimen profile are shown. Finally, Figure 5c presents results obtained after extending the specimen profile using the proposed method.

In case of the considered drop of molten steel (see Fig. 5a), contact angles are visibly below 90°, while contact angles determined from the original profile are above 90°. Adding the points to specimen profile using the introduced method improves the accuracy of contact angle determination as the determined values are below 90°.

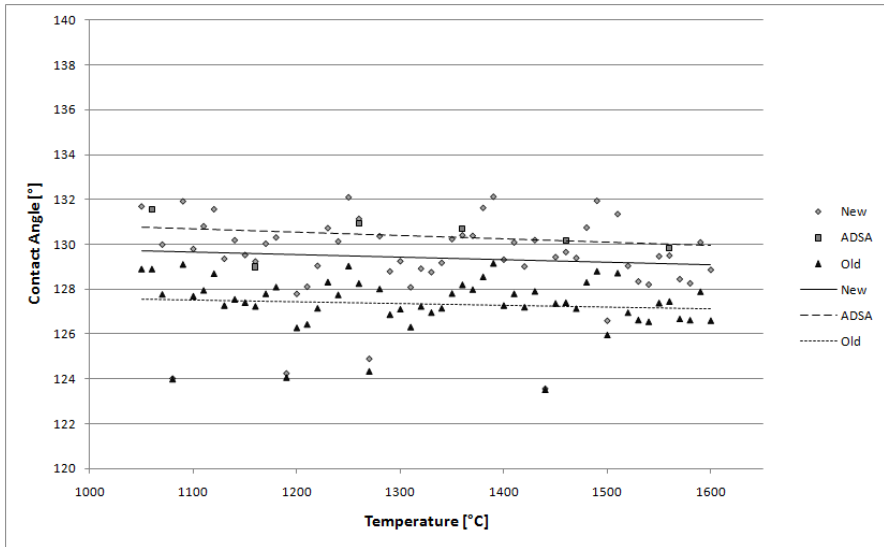


Fig. 6 Results of contact angle determination in images of copper at temperature 1050°C-1200 °C on Al_2O_3 surface.

In Figure 6 results of contact angle determination in images of copper at temperature 1050°C- 1200°C are shown. Results provided by the old and the new method are compared with results obtained using ADSA approach used as the reference.

The old method for contact angle measurements provides angles characterized by high deviation. Moreover, the determined values of contact angles are significantly smaller than the reference provided by ADSA.

The introduced method for contact angle determination significantly improves quality of the obtained results. The determined values of contact angles are very close to the reference values. Moreover, they have the same - decreasing trend.

It should be also underlined, that ADSA method used as the reference is characterized by high computational complexity and in consequence is very time consuming. The proposed method is much simpler and significantly less computationally complex. This makes it very useful in "Thermo-Wet" system as remarkably accelerate process of contact angles determination.

5 Conclusions

In this paper problem of accurate contact angle determination from images of molten heat-emitting materials was considered. The new method for contact angle determination was proposed. The method was developed for computerised system for high temperature measurements of metals and alloys.

Experimental results have proven that the introduced method provides results close to those provided by axisymmetric drop shape analysis. However, the new method is significantly less complex than ADSA what accelerates the measurements.

It should be also underlined, that although the proposed method was developed for the certain measurement system it can be successfully applied in other applications for contact angle determination.

Acknowledgement. This paper presents research sponsored by the Ministry of Science and Higher Education of Poland, as a project no. N N519 403037.

References

- [1] Mortensen, A.: Interfacial phenomena in the solidification processing of metal-matrix composites. *Material Science and Engineering A-135*, 1–11 (1991)
- [2] Hosking, M., Yost, F.G.: *The mechanics of solder alloy wetting & spreading*. Springer, Germany (1993)
- [3] Nicholas, M.G.: *Joining processes*. Springer, Germany (1998)
- [4] Deyev, G.F.: *Surface phenomena in fusion welding processes*. CRC Press, USA (2005)
- [5] Evans, J.W.: *Guide to lead-free solders: physical metallurgy and reliability*. Springer, Germany (2007)
- [6] Adamson, A.W., Gast, A.P.: *Physical chemistry of surface*. Wiley-Interscience, USA (1997)
- [7] DeGennes, P., Brochard-Wyart, F., Quere, D.: *Capillarity and wetting phenomena: drops, bubbles, pearls, waves*. Springer, Germany (2003)
- [8] Hartland, S.: *Surface and interfacial tension: measurement, theory, and applications*. CRC Press, USA (2004)
- [9] Sankowski, D., Strzecha, K., Jezewski, S.: Digital image analysis in measurement of surface tension and wettability angle. In: *Proc. IEEE Int. Conf. Modern Problems of Telecommunications, Computer Science and Engineers Training*, pp. 129–130 (2000)
- [10] Sankowski, D., Senkara, J., Strzecha, K., Jezewski, S.: Automatic investigation of surface phenomena in high temperature solid and liquid contacts. In: *Proc. IEEE Instrumentation and Measurement Technology Conference*, pp. 1397–1400 (2001)
- [11] Strzecha, K., Koszmider, T.: Drop shape analysis for measurements of surface tension and wetting angle of metals at high temperatures. In: *Proc. IEEE Int. Conf. Modern Problems of Telecommunications, Computer Science and Engineers Training*, pp. 57–59 (2008)
- [12] Fabijańska, A., Sankowski, D.: Improvement of image quality of high-temperature vision system. *Measurement Science and Technology* 20(10), 104018 (9pp) (2009)

A Fuzzy C-Means Based Color Impulse Noise Detection and Its Benefits for Color Image Filtering

Mihaela Cislariu, Mihaela Gordan, Victor Eugen Salca, and Aurel Vlaicu

Abstract. Many median filters are developed for images affected by color impulse noise. A particular approach aims to preserve fine details by noise detection followed by filtering. The color noise detection algorithms vary as principle and performance. This paper proposes a new color image filtering method from this class, which jointly applies two methods of modified fuzzy c-means clustering for the detection of noisy pixels and afterwards performs a color noise filtering on the detected pixels only. The approach shows a good noise detection performance (in terms of false acceptance and false rejection rates), and the filtering performance in terms of PSNR and details preservation is superior to other filters (including vector median filter).

1 Introduction

Noise can be systematically introduced into digital images during acquisition or transmission of images. A fundamental problem of image processing is to effectively reduce noise from a digital image while keeping its features intact. Three main types of noise exist: impulse noise, additive noise and multiplicative noise. The impulse noise is a frequent and very disturbing type of noise in digital images, usually characterized by some portion of image pixels that are corrupted, leaving the remaining pixels unchanged. In the case of grey scale images, it usually appears as black and white dots distributed over the image (therefore its more common name of "salt & pepper" noise); in color images, the effect is maybe even more disturbing and harder to remove, since the "impulses" can alter the three color channels, appearing as color impulse noise. Therefore specific algorithms must be developed for an efficient impulse noise removal in color images, which are not always straightforward generalizations of the grey scale impulse noise filtering algorithms. It is well known

Mihaela Cislariu · Mihaela Gordan · Victor Eugen Salca · Aurel Vlaicu

Technical University of Cluj-Napoca, Cluj-Napoca, Romania

e-mail: Mihaela.Cislariu, Mihaela.Gordan@com.utcluj.ro

Victor.Salca, Aurel.Vlaicu@com.utcluj.ro

that the most common filter type for the impulse noise removal in the grey scale images is the median filter. In the case of the color impulse noise, a number of methods have been proposed in the literature, such as: the adaptive scalar median filter; the median filter applied to the chromaticity in the HSI space; the median filter based on conditional ordering in the HSV space; the vector median filter [5]; the arithmetic mean filter [7]. Many filtering techniques have been proposed to remove impulse noise; however the removal of impulse noise is often accomplished at the expense of edge altering. If we apply the median filter on the entire image, some edges will also be affected by the filter. A way for solving this problem is to apply the median filter only on the noisy pixels in the image. This implies an accurate detection of the noisy pixels performed prior to filtering - which is a non-trivial task. Recently, several algorithms for the detection of noisy pixels are proposed in the literature. Some of these methods are summarized in the following. In [6] the authors propose a no reference blur metric based iterative edge preserving filtering technique. In [4] the author proposes a two-step algorithm. In the first step he identifies the possibly corrupted pixels by ordering the pixel values by rank and in the second step he decides whether the pixel under the consideration is really corrupted or not by computing the distance vector and comparing the current pixel with the minimum and maximum value of the noise free pixels. In [3] the authors propose an adaptive sub-band-based strategy in which an image is divided into low-frequency blocks and high-frequency blocks of 8×8 pixels based on their PSNR, estimated between the original brightness block and a low-pass filtered version of the block performed directly in the discrete cosine transform domain through coefficient quantization. Depending on the type of the block, a different filtering neighborhood is chosen to iteratively apply a multi-state median filter on each pixel in the block, and the PSNR between the filtered and non-filtered block is used in an iterative process to decide when to stop the iterations. In [9] the authors propose a technique that employs the switching scheme based on the impulse noise detection mechanism using so called peer groups concept. In this paper we propose an alternative approach to the color impulse noise detection process and apply afterwards a color noise filtering technique on the detected noisy pixels - e.g. the vector median filter (VMF), which is a popular technique due to its good performance in noise removal (used in a similar filtering approach in [9]). The detection of the noisy pixels prior to filtering can improve significantly the detail-preserving ability of the filter (as with any filters in general) and its computational speed, since the filtering is applied just on a sub-set of the pixels present in the image. Whereas the benefits of the noise detection step to detect the pixels on which the filtering is applied are not significant for high levels of noise density, the gain is especially remarkable for low levels of noise density, and this situation can occur in practice (as one can see in a standard test image database, TID2008 [8]). Our approach to noisy pixels detection in a color image is based on some modified version of the standard fuzzy c-means clustering [2], applied on the intensities of the pixels in the image; the modification is inspired by a noise rejection fuzzy c-means algorithm proposed in [10], where the authors include a new outlier detection and membership degree filtering stage in the standard iterative procedure of fuzzy c-means to improve the segmentation results of grey

scale medical images in the presence of noise. Inspired by this formulation, we propose to use such a modified fuzzy c-means clustering approach for noise detection by examining the amount of change of the membership values during the iterative minimization process associated with the fuzzy c-means algorithm combined with the membership degrees filtering. If a pixel is noisy, then is likely to expect that its membership degree to the clusters generated by the fuzzy c-means partition will be much different than the membership degrees of its neighbors. Furthermore, if the pixel is in a uniform region, the membership degrees of the neighbors of the noisy pixel to all the fuzzy clusters generated by the fuzzy c-means algorithm are very similar. Even if the pixel is in a region around a boundary, there can generally be only two dominant fuzzy clusters to which the pixels around the noisy pixel give high membership values. These considerations are of course valid in the case of a moderate noise density; for a very large noise density, there will be a significant difference in the membership degrees for all the pixels in the neighborhood, in which case our proposed noise detection algorithm will fail, but in such a case the efficiency of noise detection is no longer important, since most of the pixels need to be filtered anyway. Thus, in the proposed approach, we apply a noise rejection modified fuzzy c-means clustering (here is referred as NR-MFCM), we compute a cumulative membership degree amount of change in the modified fuzzy c-means iterative process from the beginning until convergence, and binary classify the pixels according to their cumulative membership degree change (denoted by CMDC) as noisy (for large values of this measure) or not noisy (for small values of this measure). This general approach can be applied on any representation of the color image. However considering that in most practical cases the images are stored and transmitted either in a luminance-chrominance encoding format or in an RGB primary color space representation format, it is likely to assume that the noise will affect independently the corresponding color space components. Therefore an independent noise detection on each such color plane would be the most straightforward approach; a pixel is classified as noisy if it was identified as noisy in at least one color plane. The description of the noise rejection modified fuzzy c-means clustering, the proposed approach, the implementation, the results and the conclusions are presented in the following.

2 The Noise Rejection Modified Fuzzy C-Means Pixels Clustering(NR-MFCM)

Let $X[H \times W]$ represent an intensity image, each pixel in a spatial location (i, j) being in this case a scalar value $x(i, j) \in \mathfrak{R}$. If on the set of all intensity values $\{x(i, j) | i=1, \dots, H, j=1, \dots, W\}$ one applies a fuzzy c-means clustering [2], based on the Euclidean distance between the intensity values of the pixels in X (without taking into account the spatial distribution of the intensity values in the image), we obtain at the end of the iterative fuzzy c-means algorithm, the optimal fuzzy partition matrix $U[C \times HW]$, showing the membership degrees of each data $x(i, j)$ to the C fuzzy classes, and the C prototypes of the classes v_k (intensity values obtained as the

centers of mass of the corresponding fuzzy classes), $k = 1, \dots, C$. For an easier description of the NR-MFCM algorithm, we consider an alternative representation of the membership degrees of the pixels in X to the C fuzzy classes, in the form of a 3D matrix $\tilde{U}[H \times W \times C]$, with $\tilde{u}(i, j, k)$ the membership degree of the pixel $x(i, j)$ to the fuzzy class k , $k = 1, 2, \dots, C$. $U_k[H \times W]$ is the " k^{th} plane" of the matrix \tilde{U} , formed by the membership degrees of the pixels in X to the fuzzy class k . The optimal values of membership degrees and fuzzy prototypes are found by the minimization of the cost function:

$$J_m(\tilde{u}, v) = \sum_{i=1}^H \sum_{j=1}^W \sum_{k=1}^C \tilde{u}^m(i, j, k) d^2(x(i, j), v_k), \quad (1)$$

where m is a parameter controlling the shape of the resulting clusters; the convergence of the optimization problem is guaranteed for $m > 1$. d is a distance measure between $x(i, j)$ and v_k - usually the Euclidian distance. The minimization of J_m is solved iteratively, starting from an initial fuzzy partition matrix U^0 or an initial set of cluster centers v_k^0 , being updated in each iteration t (until convergence) with the following expressions:

$$\tilde{u}(i, j, k) = \left(\sum_{l=1}^C \left(\frac{d(x(i, j), v_k^{(t-1)})}{d(x(i, j), v_l^{(t-1)})} \right)^{\frac{2}{m-1}} \right)^{-1} v_k^{(t)} = \frac{\sum_{i=1}^H \sum_{j=1}^W \tilde{u}^{(t-1)}(i, j, k) x(i, j)}{\sum_{i=1}^H \sum_{j=1}^W \tilde{u}^{(t-1)}(i, j, k)} \quad (2)$$

Typically, in a noise-free intensity image X , the intensities $x(i, j)$ will not change abruptly in small spatial neighborhoods and therefore also the corresponding membership degrees will be similar in these spatial neighborhoods. However when a noisy pixel (i.e. with a much different intensity than its neighbors) appears in the image, this pixel will have, at least for one of the fuzzy classes k , a membership degree much different than its neighbors to this class. Therefore, examining the 3D fuzzy membership degrees matrix of the image, $\tilde{U}[H \times W \times C]$, should allow us to identify the noisy pixels as well; if some filtering schemes (e.g. spatial averaging, similar to the standard filtering procedures from image processing) are applied on each "fuzzy plane" U_k ($k = 1, 2, \dots, C$) from \tilde{U} , the resulting fuzzy clustering will be more robust to the noise. This is a useful idea if we want to apply the fuzzy C-means algorithm for image segmentation; if we consider the opposite goal, of identifying the noisy pixels instead of just "ignoring" them in the segmentation process, we can use the same approach, since the noisy pixels are the ones whose membership degrees will change the most in the filtering procedure from their values resulting by the standard fuzzy C-means iterations. The local spatial averaging of the membership degrees may be performed after the completion of the standard fuzzy C-means iterations or during each iteration step. The later case is more suitable for noise detection, since it allows to assess better the evolution of the memberships in the noise affected areas as opposed to the evolution in the noise-free areas. Let $w_{(i, j)} [3 \times 3]$ be a spatial window centered on the location (i, j) over a membership plane for the fuzzy class k , $U_k [H \times W]$, and $A [3 \times 3]$ - a matrix of positive weights, $A = \{a(p, q) | p, q = 1, 2, 3\}$, $\sum_{p=1}^3 \sum_{q=1}^3 a(p, q) = 1$, $a(p, q) \geq 0$, $\forall p, q = 1, 2, 3$, with at least two non-zero elements. Then $\bar{u}^{(t)}(i, j, k)$, $i = 1, 2, \dots, H$; $j = 1, 2, \dots, W$; $k = 1, 2, \dots, C$ is called

the filtered membership degree of the pixel $x(i,j)$ to the fuzzy class k after the t^h iteration and is computed as:

$$\bar{u}^{(t)}(i, j, k) = \sum_{p=1}^3 \sum_{q=1}^3 \tilde{u}(i+p-2, j+q-2, k) a(p, q) \quad (3)$$

These averaged membership degrees are used to re-update the membership degrees $\tilde{u}(i,j,k)$ and the prototypes values v_k in each iteration in the NR-MFCM algorithm, which can be expressed by the following steps.

Step 1. Initialization. Define: m , C , the distance measure d , set the convergence error ε . Select an exponent controlling the influence of the average in the updating of the membership degrees, $n > 1$. Generate an initial fuzzy partition matrix $\bar{U}^0[H \times W \times C]$. Compute the initial prototype values v_k according to Equation 2. Set the iteration step $t = 0$.

Step 2. Set $t=t+1$. Update the membership degrees $\tilde{u}^{(t)}(i, j, k)$ and the prototypes $v_k^{(t)}$ according to Equation 2.

Step 3. Compute the Averaged Membership Degrees. According to Equation 3, update the membership degrees by (as proposed in [10]):

$$\hat{u}^{(t)}(i, j, k) = \frac{\tilde{u}^{(t)}(i, j, k)^m (\bar{u}^{(t)}(i, j, k))^n}{\sum_{k=1}^C (\tilde{u}^{(t)}(i, j, k)^m (\bar{u}^{(t)}(i, j, k))^n} \quad (4)$$

$i = 1, 2, \dots, H; j = 1, 2, \dots, W; k = 1, 2, \dots, C$, and copy these values in the old fuzzy membership matrix: $\bar{u}^{(t)}(i, j, k) = \hat{u}^{(t)}(i, j, k)$, $i = 1, 2, \dots, H; j = 1, 2, \dots, W; k = 1, 2, \dots, C$. Update v_k^t according to Equation 2.

Step 4. If $\max |\tilde{u}^{(t)}(i, j, k) - \tilde{u}^{(t-1)}(i, j, k)| < \varepsilon$, the convergence has been reached and the algorithm stops. Otherwise, go to Step 2.

In the updating rule from Equation 4, if no outlier is present in the current spatial window $w_{(i,j)}$, $\bar{u}^{(t)}(i, j, k) \cong \tilde{u}^{(t)}(i, j, k)$ and therefore $\hat{u}^{(t)}(i, j, k) \cong \tilde{u}^{(t)}(i, j, k)$. However, if the current pixel $x(i,j)$ is a noise pixel, its membership degree $\tilde{u}^{(t)}(i, j, k)$ may be significantly different than $\bar{u}^{(t)}(i, j, k)$ and therefore $\hat{u}^{(t)}(i, j, k)$ would change significantly, especially for large values of n . This means that $x(i,j)$ will contribute more to another class' prototype, and the prototypes v_k ($k = 1, 2, \dots, C$) will change. But since in general the number of noisy pixels is much smaller than the number of "clean" pixels, the values v_k are not likely to change significantly (unlike the membership values). Therefore, if reaching Step 2 of the iterative process again, the membership values $\tilde{u}^{(t)}(i, j, k)$ will again modify according to Equation 2 until some "steady state" is reached. The larger the difference in the intensity of a pixel $x(i,j)$ from its neighbors in $w_{(i,j)}$, the longer the convergence time and the larger the cumulative membership degrees to change (CMDCD) the C fuzzy classes will be. This observation is very suitable for the detection of impulse noise; if we define CMDC feature during the NR-MFCM algorithm for each pixel $x(i,j)$ by:

$$\Delta\tilde{u}(i, j) = \frac{1}{TC} \sum_{t=1}^T \sum_{k=1}^C |\tilde{u}^{(t)}(i, j, k) - \hat{u}^{(t)}(i, j, k)| \quad (5)$$

and we compare this value with a statistically derived threshold θ (e.g. some high value percentile) the pixels whose $\Delta\tilde{u}^{(t)}(i, j, k) > \theta$ are very likely to be noisy, whereas the others are likely to be clean. This strategy is adopted in our noise detection scheme referred briefly by the acronym FCMDC (fuzzy cumulative membership degree change), explained for a color image in the following.

3 The Proposed Fuzzy Cumulative Membership Degree Change Based Algorithm (FCMDC)

Let us assume a color image is affected by color impulse noise. The noise filtering goal is to obtain the noise-free image without altering the image details, but this task is often difficult since both the noise and the fine details in the image introduce high spatial frequency components and therefore any low pass filtering procedure (linear or non-linear) is prone to affect in some way all the high spatial frequencies in the image. Therefore a better solution is to first identify the locations of the noisy pixels and apply the filter in these locations only. As explained in the previous section, the solution we propose for the detection of the noisy pixels locations is based on the definition of a feature called CMDC(cumulative membership degrees change) during the NR-MFCM algorithm. The CMDC is extracted for each pixel, in each color plane of the color image representation, and compared to some threshold value to generate a crisp classification of the pixels as noisy or clean. We assume the image affected by color impulse noise is represented in the primary RGB color space. The proposed filtering algorithm is the following.

Each color plane of the image, i.e. $\mathbf{R}[H \times W]$, $\mathbf{G}[H \times W]$ and $\mathbf{B}[H \times W]$, is represented as an intensity image (denoted generically by \mathbf{X} in the previous section). On each of the three color planes, independently, we apply two versions of the NR-MFCM algorithm, differing in the filtering coefficients considered:

1. The first is the simple mean filter between all the values in the neighborhood centered on each pixel (i, j) in the color plane, $w_{(i, j)}$, i.e. the filter with the coefficients $a(p, q) = 1/9, \forall p, q \in \{1, 2, 3\}$. We denote the resulting NR-MFCM algorithm that employs this filter by NR-MFCM1 (in Figure 1), and the resulting CMDC features computed by NR-MFCM1 with Equation 5 by $\Delta R/G/B_{NR-MFCM1}[H \times W]$.
2. The second computes the mean of the neighbors of the central pixel (i, j) in the color plane, $w_{(i, j)}$, i.e. the filter with the coefficients $a(2, 2) = 0$, and $a(p, q) = 1/8$ for the other values of p and q . We denote the resulting NR-MFCM algorithm that employs this filter by NR-MFCM2 (in Figure 1), and the resulting features computed by NR-MFCM2 with Equation 5 by $\Delta R/G/B_{NR-MFCM2}[H \times W]$.

The use of these two filters in the NR-MFCM algorithm for noisy pixels detection was considered necessary because it can be estimated that they will give different

CMDC values when the membership degree of the middle pixel to a class k is much smaller than the memberships of its neighbors, and viceversa. One should note that the use of other filters may be good for other noise distributions and the selection of the most suitable filter is a topic that worths investigating. Each of the six resulting CDCM feature maps is converted to a binary mask, by the classification of its data through a binary thresholding. The threshold detection is done statistically, based on a set of training images, to minimize the false acceptance and false rejection rates for different images and amounts of noise. For an image-independent selection of the threshold, we estimated it as the 70% percentile of the histograms of the features maps $\Delta R/G/B_{NR-MFCM1}$ and $\Delta R/G/B_{NR-MFCM2}$. The resulting binary maps in which the black pixels represent the noise and the white ones - clean pixels, are combined by a max operation to obtain the global noise mask. Finally, only in the image locations marked as noisy by the global binary noise mask, we apply any color impulse noise filter - here we chose to apply the vector median filter (as also done by other authors, e.g. [9]). The block diagram of this method is presented in Figure 1. The number of classes C for the modified fuzzy c-means segmentation should be roughly compliant to the number of dominant color modes in the image histogram. In our experiments, we chose a value of $C = 5$. The suitable value for the parameter n in Equation 4, for this value of C , was 5 or 7.

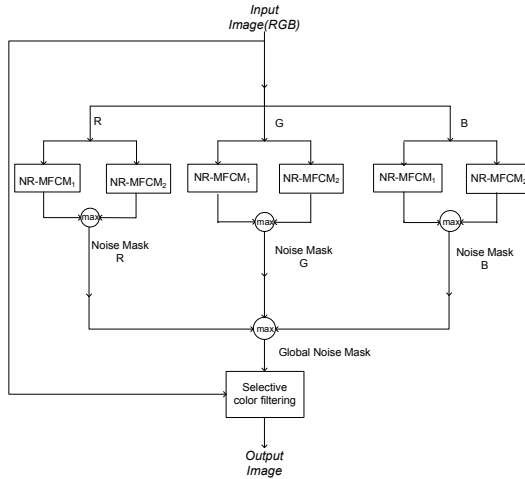


Fig. 1 The noise detection and filtering scheme

4 Implementation and Results

The performance of the proposed algorithm has been evaluated and compared with vector median filter. The algorithm was implemented in C++. A set of standard color images have been chosen for testing the performance of the filter; they were corrupted by a controlled amount of color impulse noise, at different noise levels,

using a Matlab color noise generation function. The objective quantitative measure used for the evaluation of the filtering performance is the peak signal to noise ratio (PSNR). We used the following color test images: Lenna, Basalt, Peppers and Baboon. Some experimental results, for one test image, are illustrated in Figure 2 below. One can see that both visually (in respect to the elimination of the noisy pixels) and numerically (in respect to the PSNR, summarized in Table 1), the new solution outperforms the reference vector median filter without noise detection.

Table 1 Results in PSNR for the proposed noise removal algorithm in comparison with VMF-vector median filter

Image	Noise level	Noisy	VMF	Proposed FCMDC
Baboon	0%	-	32.53	51.18
	2%	27.77	32.5	35.7
	5%	24.64	32.43	34.81
	10%	20.2	32.36	33.37
	15%	17.96	32.17	32.2
Basalt	0%	-	26.94	43.14
	2%	28.1	24.01	27.95
	5%	24.78	23.98	27.37
	10%	20.07	23.86	26.09
	15%	17.85	23.65	25.19
Lenna	0%	-	31.87	47.79
	2%	28.05	31.8	35.38
	5%	25.05	31.72	34.53
	10%	20.22	31.5	33.45
	15%	18.04	31.07	31.79
Peppers	0%	-	32.86	50.62
	2%	27.89	32.8	36.9
	5%	24.75	32.75	35.52
	10%	20.03	32.49	33.44
	15%	17.74	32.25	32.3

In terms of the noise detection performance, the FAR (showing how many non-noisy pixels were classified as noisy) and FRR (showing how many noisy pixels were "missed") were assessed on these images, for the noise levels shown in Table 1. The maximum value of FAR is around 20% and the maximum value of the FRR is around 15%, slightly better than the current results in the state of the art.

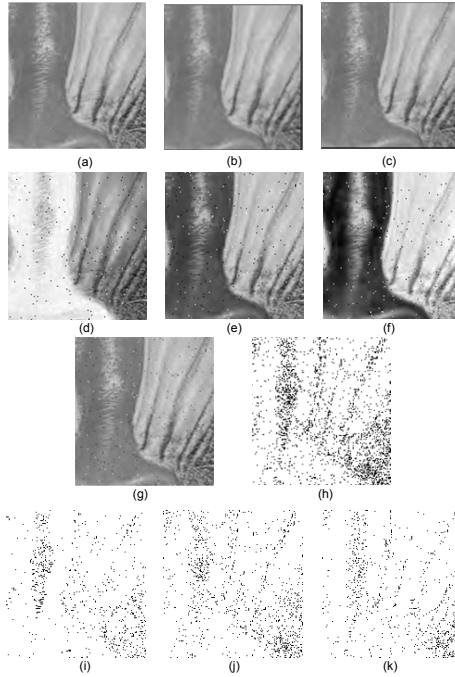


Fig. 2 (a) The original image "Baboon" detail; (b) the result for the proposed FCMDC algorithm; (c) the result of vector median filter in [1]; (d)(e)(f) the R, G and B components for the image affected by color noise; (g) the image affected by color noise; (h) Noise mask RGB; (i) Noise mask R; (j) Noise mask G; (k) Noise mask B.

5 Conclusions

In this paper a new fuzzy c-means based method for the detection of noisy pixels in color images altered by color impulse noise was proposed and its performance was assessed in conjunction with the standard vector median filter. The approach provides good noise detection results, for different images and more importantly, for different noise levels. The method is very suitable for image filtering in the case of moderate to small noise levels. Numerical measurements and visual observation show that the proposed noise detection has good performance, and its fuzzy formulation allows for extensions in terms of the interpretation of the behavior of the membership degrees changing in the noisy and clean pixels.

Acknowledgment. This work has been supported by the Doctoral Studies Support Project "PRODOC", POSDRU/6/1.5/S/5 ID 7676.

References

- [1] Andreadis, I., Louverdis, G., Chatizianagostou, S.: New fuzzy color median filter. *Journal of Intelligent and Robotics Systems* 41, 315–330 (2004)
- [2] Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)
- [3] Chen, T., Tsai, C., Chen, T.: An Intelligent Impulse Noise Detection Method by Adaptive Subband-Based Multi-State Median Filtering. In: *Proceedings of the Second International Conference on Innovative Computing, Informatio and Control*, p. 236 (2007)
- [4] Ghanekar, U.: A Novel Impulse Detector for Filtering of Highly Corrupted Images. *International Journal of Computer and Information Science and Engineering* 2(1) (2008)
- [5] Koschan, A., Abidi, M.: A comparison of median filter techniques for noise removal in color images. *University of Erlangen-Nurnberg, Institute of Computer Science* 34(15), 69–79 (2001)
- [6] Mansoor Roomi, S.M., Pandya Maheswari, T., Abhai Kumar, V.: A Detail Preserving Filter for Impulse Noise Detection and Removal. *ICGST-GVIP Journal* 7(3), 51–56 (2007)
- [7] Plataniotis, K.N., Venetsanopoulos, A.N.: *Color image processing and aplications*. Springer, Berlin (2000)
- [8] Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., Battisti, F.: TID 2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics. In: *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45 (2009)
- [9] Smolka, B., Chydzinski, A., Wojciechowski, K.: Fast Detection and Impulsive Noise Attenuation in Color Images. In: *Proceedings of the 4th International Conference on Computer Recognition Systems, CORES 2005*, pp. 459–466 (2005)
- [10] Wang, P., Wang, H.: A modified FCM algorithm for MRI brain image segmentation. In: *2008 International Seminar on Future BioMedical Information Engineering, China*, pp. 26–29 (2008)

Stereovision System for Visually Impaired

Rafal Kozik

Abstract. Nowadays the problem of BVIPs (Blind and Visually Impaired Persons) social exclusion arises to one of the major problems of modern society. It is usually framed in terms of accessibility to services like shops, theaters or cafeterias. The disability is main barrier both for fully and partially blinded to become an active members of society. However, thanks to growing progress in computer vision together with increasing power of portable devices new opportunities and solutions appear. Nearly real-time vision-based algorithms and knowledge-based systems start to help visually impaired during daily activities increasing social inclusion. New solutions for people with vision impairment are dedicated to support the user during the decision process, giving the information about obstacles located in the environment. However, in many cases this information is still not enough for blind person to have full situational awareness. Therefore the solution presented in this paper engages the stereo camera and image processing algorithms to facilitate its user with object detection and recognition mechanisms. The risk assessment based on ontology problem modeling allows to handle the risk, predict possible user's moves and provide the user with appropriate set of suggestions that will eliminate or reduce the discovered risk.

1 Introduction

Research on solutions for visually impaired still reveals that the most popular and commonly used solutions for aiding the blind people are systems based on active environment sensing ("Detection" box in Fig. 1). Such devices first generate signal (commonly ultrasonic wave) that is sent to the environment and then the

Rafal Kozik

Institute of Telecommunications, University of Technology & Life Sciences,

Kaliskiego 7, 85-796 Bydgoszcz, Poland

e-mail: rafal.kozik@upt.edu.pl

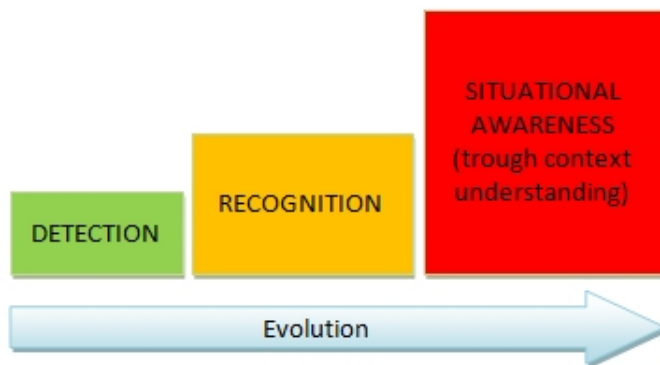


Fig. 1 Evolution of systems dedicated for visually impaired.

signal reflection is analyzed. This approach is battery consuming and vulnerable to interferences generated by another user facilitated with the same device. Moreover such solutions do not intend to replace the traditional aids such as white cane or guide dog.

However, in the last few years the research concerning the alternative solutions has been growing and introducing new designs and systems that focus on adaptive and intelligent methods utilizing stereo cues and image processing approaches. In contrary to ultrasonic-based solutions computer vision based approaches are passive, since only signal generated by the environment is analyzed (sunlight reflection, sound, etc.). Solutions (similar to one proposed in [11]) engaging the camera sensors allow to perceive the nearest obstacles. However, the users of such device still don't have enough information to tell what the detected obstacles are. Commonly the raw information about the obstacle orientation in 3D space does not allow the blind person to have a good situation awareness, therefore the object identification is crucial. It allows to predict particular object behavior (e.g. moving car) and efficiently identify threats connected with that object.

Into second group of the evolution diagram ("Recognition" box in Fig. 1) fall systems which engage pattern recognition and computer learning algorithms to recognize objects. The majority of developed algorithms serve single object detection (e.g. dedicated to find particular classes of objects), while some are aimed at identifying wide range of objects. Typical single-object detectors allow for such objects recognition as door [7], [8] pedestrians [9] or staircase [10]. However such solutions are considered by blind persons as an additional source of the information during the travel.

Into the last group in the evolution diagram ("Situational awareness" box in Fig. 1) fits solution proposed in [11]. However the MoBIC project does not adapt computer vision and focuses more on increasing the mobility of visually impaired providing information about locations. This helps the blind person to infer the surrounding threats itself.



Fig. 2 General system architecture.

The solution proposed in this paper (with vision-based object detection, recognition and ontology-based reasoning) aims at fitting into the last group.

2 System Architecture

The proposed solution has client-server architecture. There are local and remote CPUs. The local CPU is directly connected with the blind person and engages two USB cameras to perform depth estimation and obstacle segmentation (see Fig. 2). The cameras are bind together and calibrated. The depth map obtained from stereocamera allows to get the information about the position and size of the surrounding obstacles. This allows the system to plan the next step of blind person. For the nearest obstacles object recognition task is started. This task is performed on remote servers. When object is identified the ontology knowledge is queried to identify threats. Basing on visual observations (names of detected objects) system tries to identify the user localization. It allows the system to perform further reasoning like identifying the most severe threats connected with that localization. The proposed system allows the blind person to assess and to apply appropriate action plan, which will reduce or eliminate discovered threats. Finally the message (describing the threats and suggesting action plan) is generated in textual form (see "Remote CPUs" in Fig. 2) and converted to voice via TTS (Text To Speech) application.

3 Obstacles Segmentation

In order to segment the obstacles the disparity (map obtained from stereocamera) is analyzed. To obtain the real distance to obstacle (e.g. in cm) the intrinsic camera parameters have to be known. Therefore the stereo cues have to be calibrated to work with the proposed system. However the most challenging part of obstacle segmentation task is to produce disparity map.

3.1 Depth Estimation

There are different approaches to stereo matching. These can be generally divided into two groups. First group of algorithms use the local features to compute the disparity map (the product of stereo matching). The second group engages both local and global constraints, minimizing the cost function. Both of these methods have advantages and disadvantages. The local-based methods allow to obtain the disparity map in relatively short time, but its quality (accuracy) is quite poor. On the other hand the methods engaging global constraints allow to generate high accuracy of disparity maps, however not in a real time.

The algorithm for disparity map (proposed in this paper) is a hybrid one. It captures advantages of both local and global ones. Particularly it allows to estimate the depth in nearly real-time simultaneously capturing some global features like smoothness and depth discontinuities. The algorithm tries to assign each pixel in disparity map a disparity value (a label f). The energy (cost) for the assignment f is given by:

$$E(f) = E_{data}(f) + E_{smooth}(f) \quad (1)$$

The E_{data} describes the initial depth assignment cost (energy). Commonly this is a difference between luminance levels of the same pixels from left and right camera (SAD rule - Sum of Absolute Differences). The E_{smooth} describes the depth map smoothness. It can be noticed that this model does not handle the occlusion (which is considered by global approaches [2]). Tests prove that without this constraint the estimated disparity map may be very inaccurate in some conditions. Therefore to handle this issue Birchfield DP (Dynamic Programming) algorithm [3] for stereo matching was adapted to compute the E_{data} term. Instead of pixel-to-pixel matching the algorithm focuses on lines. The algorithm inserts blank spaces into one of the lines to obtain the best matching (Fig. 3). Each inserted empty space (padding) can be considered as occluded region.

It should be noticed that (in comparison to SAD) this method gives sharp edges of obstacles. It can be also noticed that some lines still have poorly estimated lines. Therefore E_{data} term has to be combined with smoothness term. In order to find the optimal solution of disparity label assignment this problem formulation can be modeled as finding MAP (Maximum A Posteriori) solution to MRF (Markov Random Field) problem. As it is stated in [6], good approximation for MAP problem is MAX Product Belief Propagation. It uses an efficient message passing technique that allows to find the solution in real-time. Examples of disparity map obtained with DPBP algorithm (Dynamic Programming + Belief Propagation) are shown in Fig. 4.

3.2 Finding Obstacles in Depth Map

Before the particular object can be located in the environment it has to be segmented and extracted from the depth map. However, this is not a trivial task since different objects may occupy different depth ranges. The main idea which solves this problem

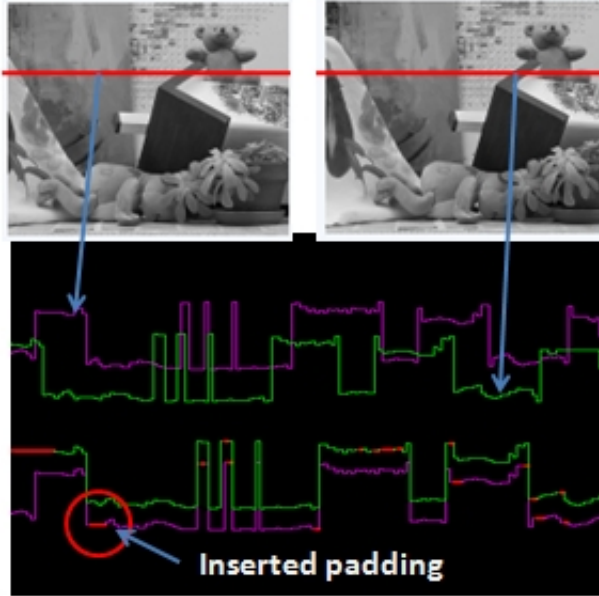


Fig. 3 Scan-line matching algorithm.

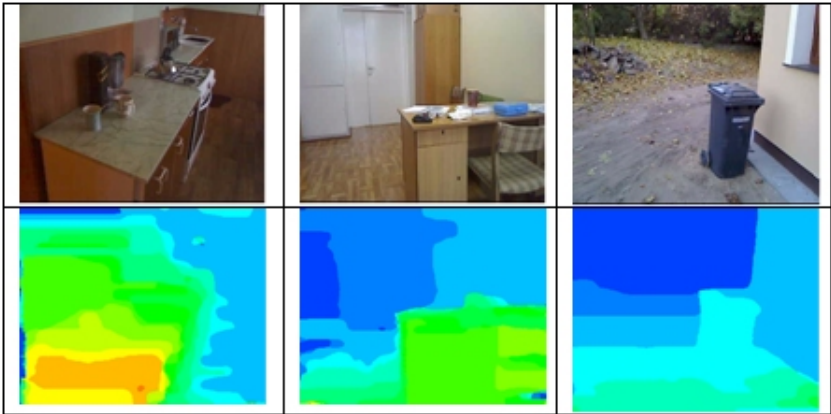


Fig. 4 Examples of depth map obtained with DPBP algorithm.

aims at generating the bird’s eye view image of the scene. This can be obtained by applying the formula Eq. 2 to each pixel of the image, where *height* indicates the depth map height, and $\text{depthMapPixEq}(x, j, y)$ is equal to 1, if pixel at position (x, j) has depth *y* and 0 otherwise.

$$birdsEyeView(x,y) = \sum_{j=0}^{height} depthMapPixEq(x,j,y) \quad (2)$$

The bird's eye view image (3D model in Fig. 6) is further segmented and each of segments is labeled to identify the width and the depth ranges of the object. Having this information the obstacle can be easily located in the original image as it is shown in Fig. 5. This region of texture is used as an input for obstacle identification.



Fig. 5 Nearest obstacles detection and labeling (red line)

4 Obstacle Identification

Since recognition is performed on remote machines huge amount of resources can be assumed. However this can be done only theoretical because whole system has to work in real-time. Therefore applied detectors and object recognition techniques have to be adjusted processing power of remote units. Currently there are different approaches to index images in large data base. However, among many following ones are adapted to work with the proposed system: SURF descriptors, BoW (Bag of words), GIST descriptors, Haar-like features. For SURF descriptors naive nearest neighbor classifier is chosen (distances between the first and the second match have to be greater than system threshold). For BoW and GIST kNN classifier is used while for Haar-like features boosted cascade of classifiers is applied. Different classifiers effectiveness and examples are presented in "Experiments" section. First tests showed that setting the stereocamera to capture in small resolution (320x240) allow to achieve real-time stereo matching. However the low resolution causes that some high resolution texture patterns can not be detected if object is too far from camera.

Therefore the solution integrating mobile phone camera with the system was proposed to increase the object recognition. In that case stereovision camera is attached to a belt tied around waist, so that only mobile phone is carried by blind person.

5 Risk Assessment

The design of the proposed system assumes that information about threats the blind person is exposed might be deducted from objects found in environment and from environment itself. Particularly, objects introducing threats are mugs, glasses, kettles or pots (not detected on time may expose blind person to serious injury). Inferring threats both about places and objects requires prior knowledge. In order to solve these complicated relations between object and to provide good semantic description the ontology notation model is introduced. The ontology focuses on risk assessment. It describes object, which has some physical properties. In example coffee mug will have property "may contain hot coffee". With that property there are connected some threats. In this example there are two of them: "burn yourself" and "spill content". Each of these threats increase the total risk the blind person is exposed to. However, for each of treats there may be some reaction that will eliminate or reduce the risk.

6 Experiments

The system effectiveness was evaluated using set of scenarios. Each one aims at modeling typical problem that blind person has to cope with. These are categorized in following groups:

- Threats identification
- Object recognition
- Obstacle avoidance

6.1 *Avoiding Collisions and Planing Next Steps*

The obstacles detected trough depth map segmentation are used to build small 3D map of surrounded environment. This allows the system to plan the next step of blind person. System analyses obstacles height and distances between them. The suggestions are generated in textual form and presented to blind person using Text-To-Speech. Examples of algorithm at work are presented in Fig. 6.

6.2 *Object Recognition*

For the experiments databases of 182 images (47 unique objects) was collected. Images were taken for three (indoor testbed) localizations: university, home and office. During the tests recognition ratio was evaluated (Fig. 7).

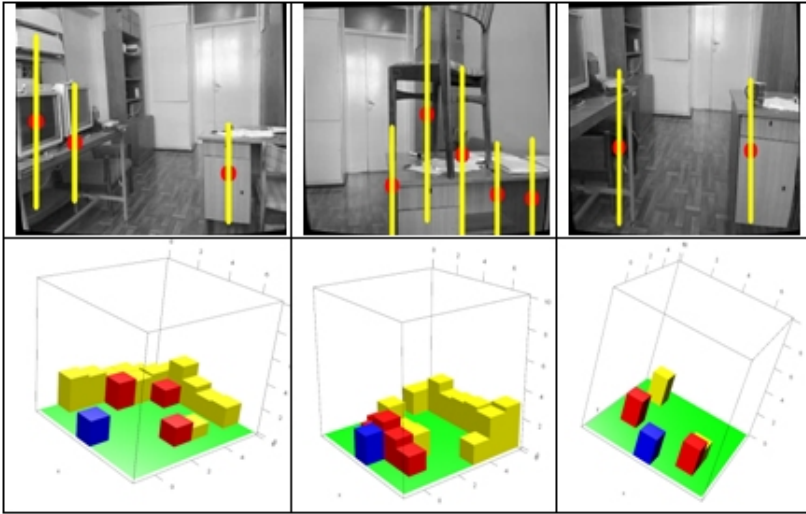


Fig. 6 Example of segmented objects (red boxes are the colliding obstacles and blue box indicates the user).

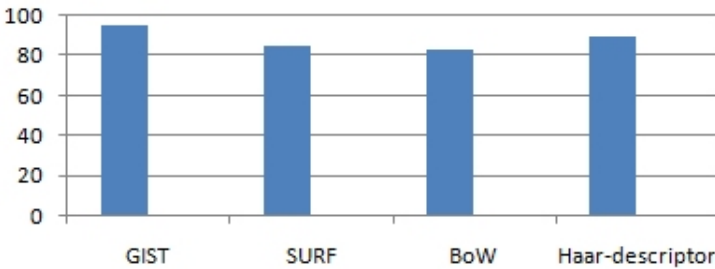


Fig. 7 Recognition ratio of adapted classifiers.

6.3 Risk Assessment

An example of risk assessment module at work is shown in Fig. 9. The left panel displays the name of most likely location. The Right panel list the objects seen by the system in last 15s. The bottom panel displays list of identified risks. Each time new object is detected and recognized by the system the ontology is queried in order to evaluate the risk. Each visual observation (recognized object) is used to estimate user localization. In the example shown in Fig. 9 the system was able to estimate the user localization with high probability (after recognizing the oven). Basing on that information system detected two threats: one coming from oven and the second one coming from prior knowledge that there is a kettle on the left side.



Fig. 8 Object recognized via mobile phone camera (additional camera sensor in the system).

Location	Seen objects
Kitchen	oven
Threats	
You can get burnt by hot oven!	
Notice! There may be a kettle on your left.	

Fig. 9 GUI panel of risk assessment module.

7 Conclusions

In this paper prototype of mobility aid system for people with vision impairment is proposed. The preliminary results are promising and show that engaging the stereovision systems together with robust object extraction algorithm and semantic description of the environment allow to assess the risk the blind person is exposed to. Further more proposed here distributed architecture allows to perform huge amount of operation in nearly real time.

References

- [1] Pelcztnski, P.: Travel Aid System for the Blind. Image Processing and Communications Challenges, 324–333 (2009)
- [2] Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions via graph cuts, In: I, pp. II: 508–515 (2001)
- [3] Birchfield, S., Tomasi, C.: Depth Discontinuities by Pixel-to-Pixel Stereo. International Journal of Computer Vision 35(3), 269–293 (1999)
- [4] Torralba, Fergus, R., Weiss, Y.: Small codes and large databases for recognition. In: CVPR (2008)
- [5] Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: Improving particular object retrieval in large scale image databases. In: CVPR (2008)

- [6] Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. In: CVPR, pp. 1: 261–1: 268 (2004)
- [7] Yang, X., Tian, Y.: Robust Door Detection in Unfamiliar Environments by Combining Edge and Corner Features
- [8] Chen, Z., Birchfield, S.: Visual Detection of Lintel-Occluded Doors from a Single Image. In: IEEE Computer Society Workshop on Visual Localization for Mobile Platforms (in association with CVPR), Anchorage, Alaska (June 2008)
- [9] Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008)
- [10] Wang, S., Wang, H.: 2D staircase detection using real adaboost. In: Proceedings of the 7th International Conference on Information, Communications and Signal Processing, Macau, China, pp. 376–380. IEEE Press, Los Alamitos (2009), ISBN: 978-1-4244-4656-8
- [11] MoBIC: an aid to increase the independent mobility of blind and elderly travellers. 2nd TIDE Congress, Paris, La Villette, April 26 - 28 (1995)

Part IV
KNOWLEDGE ACQUISITION BASED
ON REASONING METHODS

Artefactual Reasoning in a Hypergraph-Based CAD System

Mehul Bhatt, Szymon Gajek, Ewa Grabska, and Wojciech Palacz

Abstract. This paper deals with a computer system supporting conceptual floor layout and reasoning about abstract spatial artefacts. The reasoning module is able to evaluate constraints which involve immaterial entities (range spaces of sensor devices). The internal data structures of the system are hypergraph-based, but the design constraints are specified as first-order logic formulae.

1 Introduction

The first phase of the design process – the conceptual phase – is very important. Decisions taken in this phase heavily influence the later phases, and thus the whole project. In order to properly support conceptual design, a Computer Aided (Architecture) Design (CAD / CAAD) system has to be able to work with incomplete, preliminary projects, and to communicate their current state to the users. Architectural drawings are not suitable for this purpose, because in conceptual phase many properties (such as exact room dimensions, materials, etc.) have not been specified yet. Simple diagrams, similar to hand-drawn sketches, seem to be more convenient to externalize designer's ideas.

The designer has to take into account many constraints with varying degrees of importance. Some of them are dictated by physics, some are part of the building code, some are recommended as the best architectural practices. Ensuring that all of them are fulfilled is no easy task. So, there were many attempts to implement a system for automatic validation of architectural design projects. Those systems were primarily meant to be employed at the end of the design process. They were used to check if a finished project fulfills all necessary constraints [9].

Mehul Bhatt

SFB/TR 8 Spatial Cognition, University of Bremen, Germany

e-mail: bhatt@informatik.uni-bremen.de

Szymon Gajek · Ewa Grabska · Wojciech Palacz

Faculty of Physics, Astronomy and Applied Computer Science,
Jagiellonian University, Cracow, Poland

e-mail: {szymon.gajek, ewa.grabska, wojciech.palacz}@uj.edu.pl

This paper proposes a system which allows designers to create a conceptual project, and constantly monitors their progress. Any constraint violations resulting from a modification applied to the project are immediately reported.

2 The HSSDR System and the Hypergraph Model

The *Hypergraph System Supporting Design and Reasoning* (HSSDR) is a system for designing floor layouts using the top-down methodology [5, 6]. It uses a specialized visual language of diagrams to communicate with users. The diagrams can represent architectural entities such as rooms, walls, and doors.

Internally, the project is represented as a *hierarchical hypergraph*, consisting of *hyperedges* and *nodes*. In our approach there are two types of hyperedges: those representing rooms and areas, and those representing adjacency and accessibility relations. Functional areas and groups of rooms are at higher hierarchy levels, while single rooms are at the bottom. Graph nodes correspond to walls. Attributes are used to store rooms positions, as well as wall lengths, number and positions of doors, room labels, etc. A detailed description of the graph data model can be found in [7].

2.1 HSSDR as a Sketching System

A designer starts working in HSSDR by sketching the outline of an apartment. Then he divides it in two by drawing a polyline. The resulting areas are further divided, and this process continues recursively until it reaches the level of rooms.

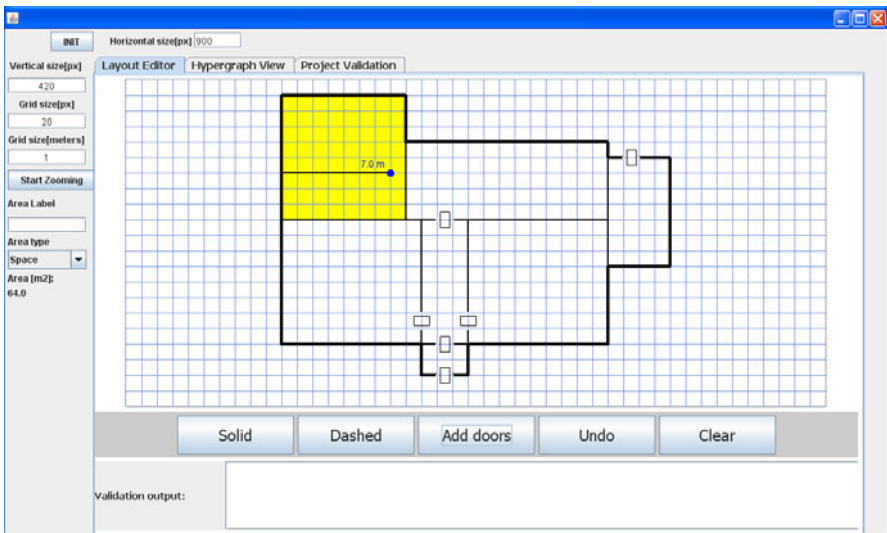


Fig. 1 A Diagram Representing a Floor Layout Being Sketched

Either during this process, or at its end, doors can be placed in the walls. An example of a floor layout being sketched is presented in Fig. 11.

Every splitting operation results in creation of two new hyperedges corresponding to two new rooms/areas. They are attached as children of the hyperedge corresponding to the original, undivided area. Thanks to that, the graph model represents not only the current state of a layout, but also its development history.

2.2 HSSDR as a Reasoning System

HSSDR was conceived as a design system aimed at assisting its users by checking if the created design fulfills specified constraints. In the beginning, the requirement constraint checking engine was dealing directly with a graph-theoretic representation and corresponding computational algorithms¹. Subsequently, we have noted that majority of the tests can be describe directly using a symbolic specification, e.g., such as the language of the first-order logic. Declaratively specifying tests in the form of the first-order logic formulae has many advantages over implementing them using procedural or object-oriented languages, primarily among them being easy specification and readability, modifiability, and quite often, a direct correspondence the manner of their conceptualisation by experts, designers, customers, and even design guidelines.

For example, a test which forbids adjacency between a kitchen and a bedroom can be written as follows:

```
failure_msg "a kitchen should not be adjacent to a bedroom"
forall x, y in Rooms : type(x) = "Bedroom" and
    type(y) = "Kitchen" => not adjacent(x, y);
```

The current version of HSSDR includes modules (see Fig. 12), which can load such formulae from external text files and evaluate them. The system lets users choose which tests they want to apply. It also allows them to write their own tests.

Valuation of a first-order logic formula can be done only in context of a given relational structure [4]. The structure provides semantic meaning for predicates, functions, and constants used in the formula. It also provides an universe of individuals over which quantifiers range.

HSSDR constructs a relational structure on the basis of the graph model. For every hyperedge representing a room it creates a corresponding individual in the structure; a binary relation named “adjacent” is constructed so that it is true only for individuals corresponding to adjacent rooms in the graph; and so on. Any modification applied to the model is immediately reflected in the structure.

This structure is used to evaluate test suites. Results of evaluation and failure messages are presented to the designer.

An example of tests checking for conformity with the *Polish Fire Code* was presented in [7]. The test classifies rooms and doors (some of them are safe areas,

¹ Tests were implemented as Java classes, and acquired the necessary data directly from the underlying graph model.

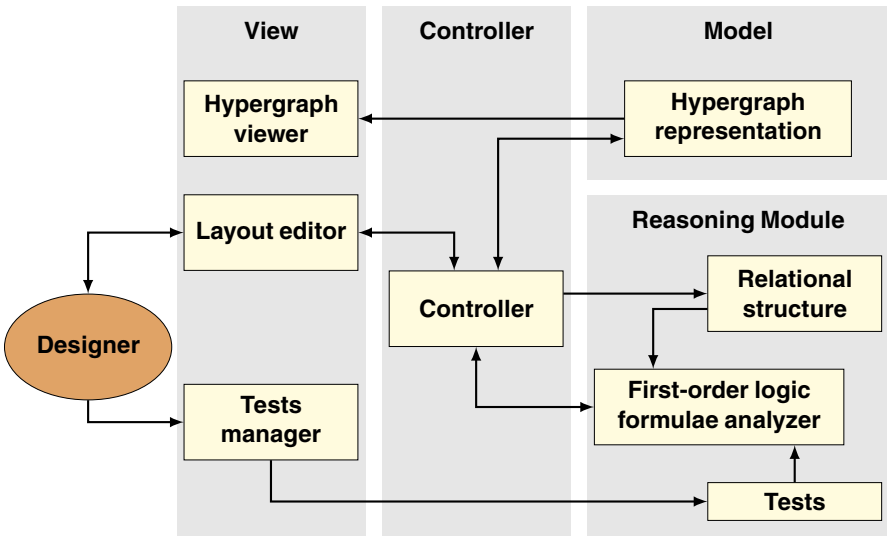


Fig. 2 HSSDR System Architecture

some others can be used as evacuation routes), then calculates the distance from every room to the nearest safe area. The fire code specifies an upper bound on this distance – going over it means that the project is not conformant.

Nearly all constraints and auxiliary definitions were specified as the first-order logic formulae. The function which calculated the distance between two doors in a room was the only exception – it had to be implemented in Java.

3 Constraints on Spatial Artefacts

Semantic descriptions of designs and their high-level requirements at an early design stage acquires real significance when functional constraints are among strictly spatial entities as well as abstract *spatial artefacts* [2]. For instance, although it is possible to model the spatial layout of an environment at a fine-grained level in conventional CAD tools, it is not possible to model *spatial artefacts* such as the *range space* of a sensory device (e.g., camera, motion sensor, view-point of an agent), which is not strictly a spatial entity in the form of having a material existence, but needs to be treated as such nevertheless.

The graph model used in HSSDR, as well as models employed by many other CAD systems, do not represent abstract *spatial artefacts*. For example, they do not represent the *operational space* of a door (the space occupied when the door is being opened), or its *functional space* (the space required by a human which wants to pass through the door). In general, architectural working designs only contain physical entities. Therefore, it becomes impossible for a designer to model constraints

involving spatial artefacts at the design level. For instance, consider the following constraint: ‘*the motion-sensor should be placed such that the door connecting room A and room B is always within the sensor’s range space*’. In order to model such constraints, rich artefactual characterizations and constraints are necessary.

3.1 Spatial Artefacts

Inspired by [2, 3], we have evaluated the feasibility of implementing tests which take into account *spatial artefacts*. Here, the following spatial artefacts may be identified; more detailed explanation of these issues can be found in [2].

- the *operational space* denotes the region of space that an object requires to perform its intrinsic function that characterizes its utility or purpose
- the *functional space* of an object denotes the region of space within which an agent must be located to manipulate or physically interact with a given object
- the *range space* denotes the region of space that lies within the scope of a sensory device such as a motion or temperature sensor, or any other entity capable of visual perception. Range space may be further classified into other categories, such as *observational space* (e.g., to model the concept of the *isovist* [1])

As it turns out, the architecture of HSSDR is naturally suited to incorporating these artefactual extensions. We discuss this in the rest of the paper with concrete examples.

3.2 Artefactual Constraints with HSSDR

Constraints are specified as logic formulae, and their value is determined on the basis of contents of the relational structure. This structure was so far constructed as a direct reflection of the graph model, and represented the same objects and relations as this model. But it can be extended, so that it contains additional information indirectly derived from the graph model. This additional information can represent properties of abstract spatial artefacts.

We validate this approach by implementing a set of tests dealing with *range space* of sensors. There are several types of sensors used in buildings: motion detectors, CCTV cameras, smoke detectors, etc.

It is possible to devise many different conditions, which could be used to constrain sensor placement. For example, we may require all paths leading from external doors to important areas (office rooms, etc.) to be monitored. So, we start by defining a one-argument predicate specifying which rooms are important:

```
SecuredArea(r) <=> Rooms(r) and (type(r) = "Office"
    or type(r) = "Living_Room" or type(r) = "Garage");
```

HSSDR provides a core predicate, *isPathWatched*, which checks if a path between two doors in the same room is inside the range space of a sensor. We define a negated version of it:

```
UnwatchedPath(d1,d2) <=> d1 != d2 and Doors(d1) and Doors(d2)
and (exists r in Rooms: doorsInRoom(d1,r) and
doorsInRoom(d2,r) and not isPathWatched(d1,d2));
```

Next, induction is applied to compute the transitive closure of *UnwatchedPath*. The induction goes up only to 99, instead of infinity, but this should be enough for all architectural projects of moderate size.

```
UnwatchedPathInduction(d1,d2,n) <=>
(n=1 and UnwatchedPath(d1,d2)) or
(n<>1 and UnwatchedPathInduction(d1, d2, n-1)) or
(n<>1 and exists d in Doors:
UnwatchedPathInduction(d1,d,n-1) and UnwatchedPath(d,d2));
UnwatchedPathPlus(d1,d2) <=> UnwatchedPathInduction(d1,d2,99);
```

Finally, the test itself is specified:

```
failure_msg "path leading to secured area is not surveilled"
not exists d1, d2 in Doors: exists r in Rooms:
SecuredArea(r) and doorsInRoom(d2,r) and
UnwatchedPathPlus(d1,d2);
```

If cameras are used instead of infrared motion detectors, then we may need an additional constraint: cameras should not be placed in private areas, and they should not observe doors leading to private areas (bedrooms, bathrooms etc). These conditions can be specified as a set of logic formulae, too.

3.3 Implementation

Implementing this new set of tests required modifications of HSSDR.

1. The graph model had to be extended so that it could represent cameras and their positions. It gained additional nodes representing sensors (these nodes are hierarchically nested in component hyperedges representing rooms).
2. A set of *Sensor* objects, relations (*sensorInRoom*, *isPathWatched*) and functions (*surveilledDoors*) had to be added to the relational structure.
3. The user interface has been modified to allow placing sensors on the layout. For each sensor the user specifies its location and direction in which it is pointing. Range and angular extent of all sensors are parametrized.

After placing a sensor it is shown as a light gray transparent circular segment. Direction, range and extent of the shape correspond to the features processed by reasoning modules of HSSDR, however visualization is not completely accurate, circular segments are not clipped when sensor's range is limited by a wall. On the other hand, on the level of reasoning modules area covered by a sensor is computed accurately.

Two figures illustrate the process of checking design constraints. For the layout presented in Fig. 3, the test concerning the security of important areas has been

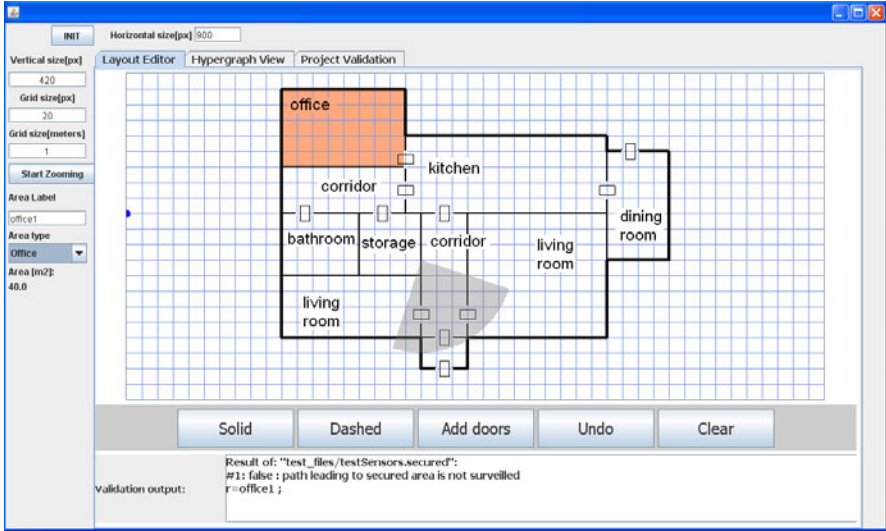


Fig. 3 A Floor Layout Not Meeting Design Constraints

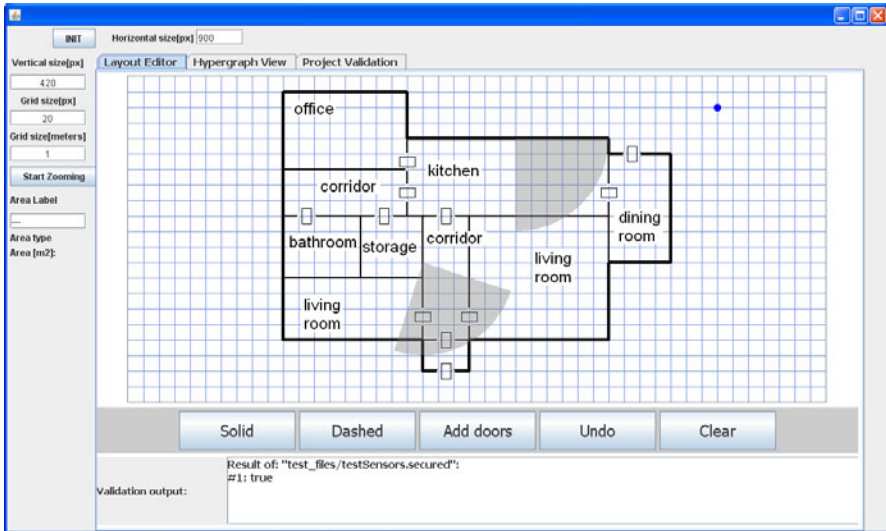


Fig. 4 After Adding Another Sensor the Layout Passes the Test

enabled. The test fails for this layout, because there is a path leading from back doors to the office that is not surveilled.

After placing another sensor in the kitchen (Fig. 4) the layout passes the test. All paths are surveilled.

4 Conclusions

This paper presents a system for conceptual design of floor layouts which can reason about spatial artefacts, and evaluate constraints which refer to them (e.g. “a bank vault door must be within a camera’s range space”). The system internally represents floor layouts as hierarchical hypergraphs. The reasoning module takes such a graph model, and derives a relational structure from it; the structure is then used to evaluate first-order logic formulae which specify design constraints.

The separation between project models and design constraints made implementing spatial artefacts easy. The reasoning module simply needs to generate structures which contain objects, relations and functions pertaining to them.

During this work we found that graph models used by HSSDR are problematic, when it comes to representing low-level layout elements (doors, sensors). We plan to replace them with models based on definitions presented in [8].

Another line of future extensions to the HSSDR system is the inclusion of formal qualitative spatial abstraction and spatial computing techniques, and semantic (ontological) representation and reasoning capabilities as proposed in [2]. This will enable us to represent the general spatial constraints and also the artefactual constraints with respect to the formal semantics of the qualitative spatial formalisms in the context of industrially conformant CAD models.

Acknowledgements. Mehul Bhatt acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) / German Research Foundation in the Transregional Collaborative Research Center *SFB/TR 8 Spatial Cognition*, Project [DesignSpace].

References

- [1] Benedikt, M.L.: To take hold of space: isovists and isovist fields. *Environment and Planning B: Planning and Design* 6(1), 47–65 (1979)
- [2] Bhatt, M., Freksa, C.: Spatial computing for design: An artificial intelligence perspective. In: *Visual and Spatial Reasoning for Design Creativity (SDC 2010)* (2011)
- [3] Bhatt, M., Dylla, F., Hois, J.: Spatio-Terminological Inference for the Design of Ambient Environments. In: Hornsby, K.S., Claramunt, C., Denis, M., Ligozat, G. (eds.) *COSIT 2009*. LNCS, vol. 5756, pp. 371–391. Springer, Heidelberg (2009)
- [4] Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: *Reasoning About Knowledge*. MIT Press, Cambridge (2003)
- [5] Grabska, E., Lembas, J., Łachwa, A., Ślusarczyk, G., Grzesiak-Kopec, K.: Hierarchical layout hypergraph operations and diagrammatic reasoning. *Machine Graphics & Vision* 16(1/2), 23–38 (2008)
- [6] Grabska, E., Ślusarczyk, G., Lan Le, T.: Visual design and reasoning with the use of hypergraph transformations. In: *Proceedings of GT-VMT 2008*. *Electronic Communications of the EASST*, vol. 10, pp. 305–318 (2008b)
- [7] Grabska, E., Borkowski, A., Palacz, W., Gajek, Sz.: Hypergraph system supporting design and reasoning. In: Huhnt, W. (ed.) *Computing in Engineering EG-ICE Conference 2009*, pp. 134–141 (2009)
- [8] Palacz, W.: Algebraic hierarchical graph transformation. *JCSS* 68(3), 497–520 (2004)
- [9] Yurchyshyna, A.: Modélisation du contrôle de conformité en construction: une approche ontologique. PhD thesis, Université de Nice-Sophia Antipolis (2009)

A Shape Description Language for Osteophytes Detection in Upper Surfaces of the Metacarpophalangeal Joints

Marzena Bielecka, Andrzej Bielecki, Mariusz Korkosz, Marek Skomorowski, Kamila Sprężak, Wadim Wojciechowski, and Bartosz Zieliński

Abstract. In this paper, the successive module of the computer system for hand diseases diagnosis support is presented. Based on the upper surface borders of the metacarpophalangeal joints, syntactic description is constructed. Then, it is analysed to verify if considered border contains osteophytes. As has been presented, shape description introduced in this paper can be used to distinguish contour of the upper surface with and without osteophytes. This kind of diversification is required to build an intelligent system for joint diseases diagnosis.

1 Introduction

Hand radiographs are the source of important clinical information in rheumatology. One of the most significant radiological features are osteophytes, which

Marzena Bielecka

Department of Geoinformatics and Applied Computer Science,
Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, Mickiewicza 30, 30-059 Cracow, Poland
e-mail: bielecka@agh.edu.pl

Andrzej Bielecki · Marek Skomorowski · Bartosz Zieliński
Institute of Computer Science, Jagiellonian University,
Łojasiewicza 6, 30-348 Cracow, Poland
e-mail: [bielecki,skomorowski}@ii.uj.edu.pl](mailto:{bielecki,skomorowski}@ii.uj.edu.pl),
bartosz.zielinski@uj.edu.pl

Mariusz Korkosz

Division of Rheumatology, Department of Internal Medicine and Gerontology,
Jagiellonian University Hospital, Śniadeckich 10, 31-531 Cracow, Poland
e-mail: mariuszk@mp.pl

Kamila Sprężak · Wadim Wojciechowski
Department of Radiology, Jagiellonian University,
Collegium Medicum, Kopernika 19, 31-501 Cracow, Poland
e-mail: kamsprezak@onet.eu, wadim@mp.pl

occur in non-inflammatory diseases, but do not occur in inflammatory diseases [6]. Detecting and tracing pathological changes is a crucial point in medical diagnosis. It is essential to distinguish between inflammatory diseases, which can be fatal, and non-inflammatory diseases, which are relatively harmless and occur in the majority of people aged around 65 [15].

Therefore, the possibility of such analysis performed by a computer system, which enables not only the detection of minimal changes in X-ray picture, but also their medical interpretation is a key point for diagnosis support. Studies concerning possibilities of such systems implementation are topic of numerous publications [5, 8, 10, 11, 12, 13, 16].

This paper is a continuation of studies described in [1, 2, 3, 4, 17, 18], concerning automatic hand radiographs analysis. In previous papers the preprocessing and joint location algorithms were introduced [4, 17, 18]. Based on obtained locations, the algorithm for identifying the borders of the upper and lower joint surfaces was proposed [3]. The preliminary analysis of such borders due to erosions detection, using Jakubowski shape description methodology [7] is described in [1, 2].

In this paper, the successive module of the computer system for hand diseases diagnosis support is introduced. Based on the upper surface borders of the metacarpophalangeal joint, syntactic description is constructed. Then, it is analysed to verify if considered border contains osteophytes.

2 A Shape Description Language

Let us consider the alphabet Σ , which consists of two types of symbols, arcs (set A) and angles (set B), defined as follows:

$$\Sigma = A \cup B, \quad (1)$$

$$A = \{[\alpha] : \alpha \in \bigcup_{a=-12}^{12} a \cdot 30\}, \quad (2)$$

$$B = \{\beta : \beta \in \bigcup_{b=-6}^6 b \cdot 30\}, \quad (3)$$

where $[\alpha]$ represents the arcs, called primitives, which start in any points called tails and end in any other points called heads. Moreover, that kind of arcs have start angles (α^{start}) and end angles (α^{end}), such that $\alpha^{end} - \alpha^{start} \in [\alpha - 15, \alpha + 15)$. The example of the arc represented by [60] is shown in Fig 1. Its tail is located near point (7.7, -6.4) and its head is located in point (0, -10). Its start angle equals 130° and end angle equals 180° .

If α is smaller than 0, arc is considered as concave, otherwise it is convex. Examples of convex arcs, represented by [60] are shown in Fig 2. Concave arcs can be simply obtained, by switching heads and tails of those primitives.

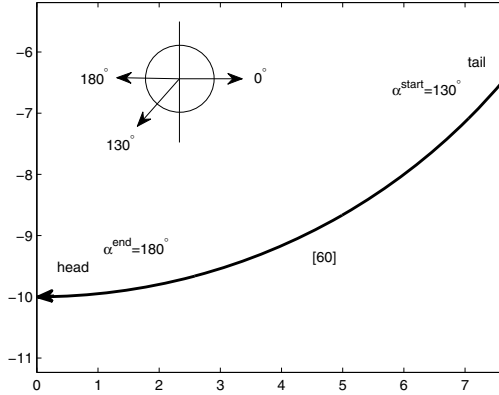


Fig. 1 Example of the primitive.

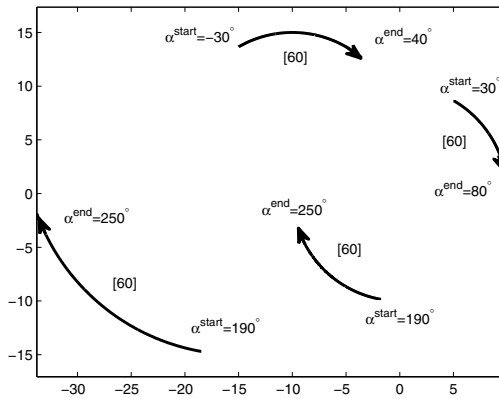


Fig. 2 Examples of arcs represented by [60].

The shape description language L consists of words k , called contours, formed alternatively from arcs $[\alpha]$ and angles β between them:

$$k \in L \Leftrightarrow k = [\alpha_1]\beta_1[\alpha_2]\beta_2\dots\beta_n[\alpha_{n+1}] \tag{4}$$

The interpretation of $[\alpha_1]\beta_1[\alpha_2]$ is as follows. Two arcs, represented respectively by $[\alpha_1]$ and $[\alpha_2]$ symbols, are connected with each other in such a way, that $head[\alpha_1] = tail[\alpha_2]$. β_1 equals $\alpha_2^{start} - \alpha_1^{end}$ rounded to the nearest angle from set B . Example contour described by $[60]90[-120]$ is shown in Fig 3b.

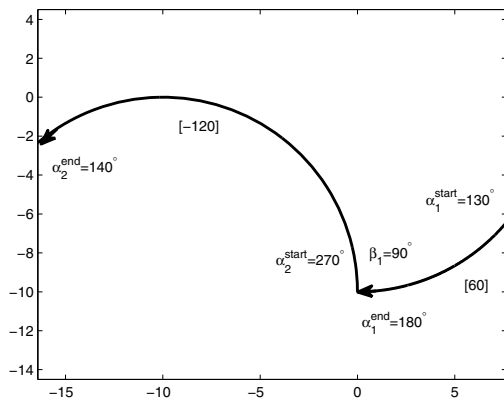


Fig. 3 Contour described by $[60]90[-120]$.

3 Contour from the Border

This approach, which is resistant to rotation, is based on the fact that numerically a tangent at the point x_n is calculated using a secant: $f'(x_n) = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$, where x_n and x_{n-1} are successive points in a numerical procedure. According to the definition of a tangent, the closer to each other x_n, x_{n-1} points are, the better approximation of a derivative is given by the above formula [9].

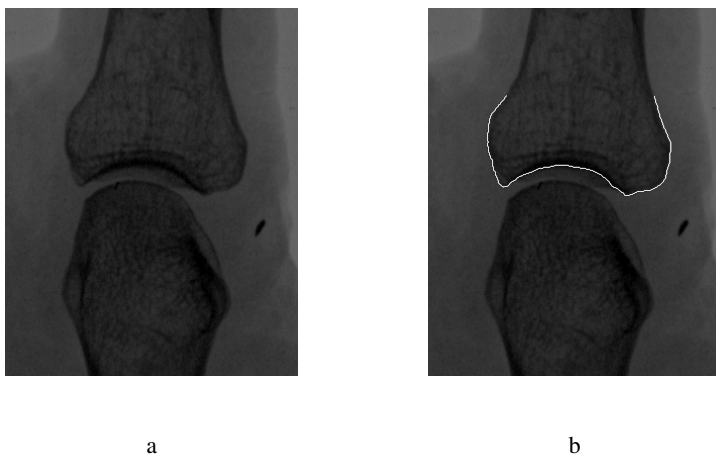


Fig. 4 Example of the healthy metacarpophalangeal joint (a) and its border (b).

Let us consider healthy metacarpophalangeal joint and its border presented in Fig 4a and Fig 4b, respectively. At the beginning, to obtain contour, the border is divided into groups of n successive points $\{(X_{P_i}, Y_{P_i})\}_{i=1}^n$. Middle point of each group is then computed as $(\frac{\sum_{i=1}^n X_{P_i}}{n}, \frac{\sum_{i=1}^n Y_{P_i}}{n})$ - white dots in Fig 5a. For each group G_i , two middle points of adjacent groups (G_{i-1} and G_{i+1}) are taken and line passes trough those two points is considered. Following [9], this line is a tangent to the middle point of the group G_i - white line in Fig 5a. Angles of tangents at successive middle points are presented in Fig 5b. Function created for such values (white thin fragments of lines in Fig 5b) is analysed using linear least squares method, to obtain fragments of the lines (grey thick fragments of lines in Fig 5b). Each line fragment contains at least m successive values for m successive middle points. Moreover, the difference between each approximated and accurate value has to be smaller than $\Delta\alpha$.

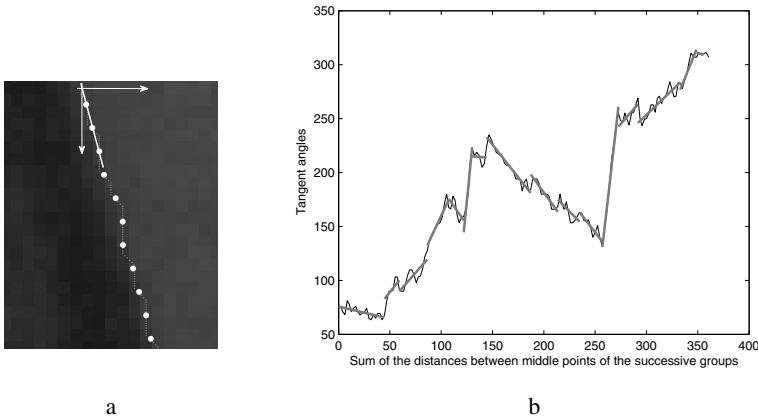


Fig. 5 Middle points of the groups created for initial fragment of border from Fig 4b (a) and tangent angles function of all middle points (b).

Further analysis is based on the fact that in case of a circle, tangent angles increase linearly. Therefore, values for the start and end of the successive line fragments can be considered as start angle and end angle of the arc (primitive), and can be described as $[\alpha_i^{start}, \alpha_i^{end}]_{i=1}^k$. Three first line fragments from Fig 5b will be then described as: [75.7523, 65.5752], [82.9490, 100.1008], [89.9691, 119.3453].

Those pairs are used to create contour in such a way, that the nearest arc from set A is taken for $\alpha_i^{end} - \alpha_i^{start}$ and the nearest angle from set B is taken for $\alpha_{i+1}^{start} - \alpha_i^{end}$. Thus, in case of considered example, the contour is rounded version of $[-10.1771]17.3738[17.1518] - 10.1317[29.3762]$, ie [0]30[30]0[30] (Fig 6b). The whole contour shown in Fig 6a equals:

$$[0]30[30]0[30]0[30]0[-30]0[90]0[0]30[-60]30[-30]0[-30]0[-30]0[120] - 30... \\ \dots[30] - 30[30]0[30]0[0].$$

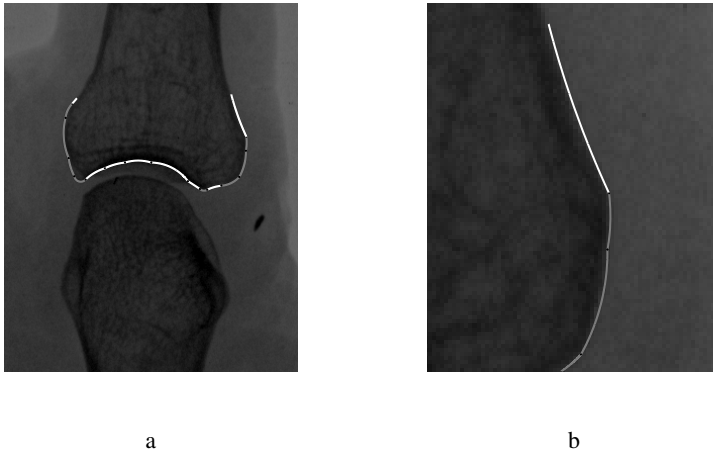


Fig. 6 Contour of the border from Fig 4b and its enlarged initial fragment (b). Gray and white curves represent convex and concave primitives, respectively.

4 Contours of the Borders with Osteophytes

Contours of the borders with osteophytes contain fragments which do not occur in case of the healthy joint contours. Those fragments can be described as patterns, which can be used to verify, if there are osteophytes in other contours.

Let us consider two metacarpophalangeal joints with osteophytes presented in Fig 7a and Fig 7b, with contours equal:

$$[0]0[30]0[30] - 30[90]0[0] - 30[0]30[-30]0[-30] - \mathbf{30[0]0[30]30[60]0[-30]-30} \dots$$

$$\dots[60]0[0]30[0]$$

and

$$[-30]0[60]30[-30]0[30]0[0] - 30[30]0[-30]0[0]0[-30]0[0]0[90]0[-30]0[0]0 \dots$$

$$\dots[30] - 30[30]0[0],$$

respectively. It can be noticed, that osteophytes are described by the fragment in bold. However, in case of the second syntactic description, first arc ([0]) can be replaced by arcs [-30], [-60] or [-90] (this arc can be longer, what results in a bigger difference between the end angle and the start angle). Therefore, it is replaced by [(0 - 30 | - 60 | - 90)], when considered as pattern.

Patterns for all osteophytes of the examined metacarpophalangeal joints are shown in Tab 1. It can be noticed, that two of them occur in more than one affected contours (column Quantity).



Fig. 7 Contours of the metacarpophalangeal joints with osteophytes.

Table 1 Patterns of the contour fragments with osteophytes, where [-] means $[(0|-30|-60|-90)]$.

Number	Patterns of the contour fragments with osteophytes	Quantity
1	$-30[0]0[30]30[60]0[-30]-30$	1
2	$[-]0[90]0[-30]0$	1
3	$[-]-30[150]30[-90]0$	1
4	$0[0]-60[180]0[-30]0$	1
5	$30[-60]0[60]60[-]$	1
6	$[-]-30[0]0[120]0[-]$	2
7	$[-]0[90]0[60]-30[-30]30$	1
8	$-30[-60]30[60]-30[30]60[-]$	1
9	$0[-60]0[120]0[0]0[60]-30[0]-30$	1
10	$[-]0[90]0[-30]30$	2

5 Experiments

In total, 24 metacarpophalangeal joints of index and middle fingers were analysed, derived from 12 images (for 6 patients) included in the test set, acquired through the offices of the University Hospital in Cracow, Poland. Half of the tested metacarpophalangeal joints presented osteophytes in upper surfaces.

Borders marked by the expert were used as the basis of the analysis, to prevent the error caused by possible incorrect automatic identification [3]. An experienced radiologist examined all upper surfaces of the metacarpophalangeal joints from the test set and manually marked borders using tablet Dell XT2. Images were presented

with the help of a graphical user interface and the radiologist was able to draw or correct the contours, as well as zoom in or zoom out the image. Moreover, the radiologist was able to mark the osteophytes.

Parameters adopted to convert border to contour was as follow: $n = 10$, $m = 3$ and $\Delta\alpha = 10^\circ$.

Contours of all joints were analysed, to check if they contain patterns from Tab. 1. Naturally, all contours with osteophytes matched those patterns. However, what is more important, all healthy joints did not match any of those patterns. This suggests that presented set can be used as a basis of complete set of patterns.

6 Concluding Remarks

As has been presented the shape description language introduced in this paper can be used to distinguish contours of the upper surface with and without osteophytes. This kind of diversification is required to build an intelligent system for joint diseases diagnosis.

Presented approach is resistant to rotation and uses rounded primitives. Thus it is better suited to bones which are rounded, than section like primitives. Despite the fact that this approach in some points is similar to Jakubowski shape description [7] and Shaw shape description [14], it still differs in several important aspects, what can be noticed in section 2.

This paper shows results for 24 joints. However, more than 200 joints are being actually analysed by a radiologist, providing authors with the possibility to verify their hypothesis on a more numerous set of examples in the future research.

To recapitulate, the final system will be the hierarchical one, with the following levels (starting from the lowest to highest level): preprocessing and joint location [4, 17, 18], identifying the borders [3], contour shape description analysis and the expert system to diagnose rheumatoid diseases. This kind of system will serve as an aid in radiological diagnosis of the hand radiographs.

References

- [1] Bielecka, M., Bielecki, A., Korkosz, M., Skomorowski, M., Wojciechowski, W., Zieliński, B.: Application of shape description methodology to hand radiographs interpretation. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010. LNCS, vol. 6374, pp. 11–18. Springer, Heidelberg (2010)
- [2] Bielecka, M., Skomorowski, M., Zieliński, B.: A fuzzy shape descriptor and inference by fuzzy relaxation with application to description of bones contours at hand radiographs. In: Kolehmainen, M., Toivanen, P., Beliczynski, B. (eds.) ICANNGA 2009. LNCS, vol. 5495, pp. 469–478. Springer, Heidelberg (2009)
- [3] Bielecki, A., Korkosz, M., Wojciechowski, W., Zieliński, B.: Identifying the borders of the upper and lower metacarpophalangeal joint surfaces on hand radiographs. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010. LNCS, vol. 6113, pp. 589–596. Springer, Heidelberg (2010)
- [4] Bielecki, A., Korkosz, M., Zieliński, B.: Pattern Recognition 41(12), 3786–3798 (2008)

- [5] Böttcher, J., Pfeil, A., Rosholm, A., Schäfer, M.L., Malich, A., Petrovitch, A., Seidl, B., Lehmann, G., Mentzel, H.J., Hein, G., Wolf, G., Kaiser, W.A.: *Journal of Digital Imaging* 19(3), 279–288 (2006)
- [6] Greenspan, A.: *Orthopedic Imaging: A Practical Approach*. Lippincott Williams & Wilkins, Philadelphia (2004)
- [7] Jakubowski, R.: *IEEE Transactions on Systems, Man and Cybernetics* 15(5), 642–651 (1985)
- [8] Kauffman, J.A., Slump, C.H., Bernelot Moens, H.J.: *Proceedings of the SPIE*, vol. 5747, pp. 1571–1581 (2005)
- [9] Kincaid, D., Cheney, W.: *Numerical Analysis. Mathematics of Scientific Computing*. Thomson Learning Inc., Wadsworth (2002)
- [10] van 't Klooster, R., Hendriks, E.A., Watt, I., Kloppenburg, M., Reiber, J.H.C., Stoel, B.C.: *Osteoarthritis and Cartilage* 16(1), 18–25 (2008)
- [11] Lehmann, T.M., Beier, D., Seidl, T., Thies, C.: *Proceedings of the SPIE*, vol. 5747, pp. 546–555 (2005)
- [12] Pfeil, A., Böttcher, J., Schäfer, M.L., Seidl, B.E., Schmidt, M., Petrovitch, A., Heyne, J.P., Lehmann, G., Oelzner, P., Hein, G., Wolf, G., Kaiser, W.A.: *Journal of Digital Imaging* 21(1), 104–112 (2008)
- [13] Sharp, J., Gardner, J., Bennett, E.: *Arthritis & Rheumatism* 43(6), 1378–1386 (2000)
- [14] Shaw, A.C.: *Information and Control* 14(1), 9–52 (1969)
- [15] Szczeklik, A., Zimmermann-Górska, I.: *Injury Disease. Medycyna Praktyczna*, Warszawa (2006) (in Polish)
- [16] Tadeusiewicz, R., Ogiela, M.R.: *International Journal of Applied Mathematics and Computer Science* 15(2), 305–312 (2005)
- [17] Zieliński, B.: *Schedae Informaticae* 16, 47–67 (2007)
- [18] Zieliński, B.: *Schedae Informaticae* 17(18), 45–61 (2009)

Using Graph Mining Approach to Automatic Reasoning in Design Support Systems

Barbara Strug

Abstract. In this paper an improved method of using frequent patterns to gain knowledge about the quality of new designs in design systems is presented. The hypergraph-based representation is used as an underlying structure for design representation. The application of this method to automatic evaluation of designs is also presented. Two different methods of reducing size of the set of frequent patterns are described. Experimental results are also presented.

1 Introduction

Graphs are a very useful way of representing complex objects in different domains of computer science [24]. They are used in engineering, system modeling and testing, bioinformatics, chemistry and other domains of science to represent objects and the relations between them or their parts [2, 3, 28].

In this approach hypergraphs [14], which enable us to express multi-argument relations between elements, are used. While in ordinary graphs edges always connect only two nodes, hypergraphs are composed of a set of nodes and a collection of hyperedges with different numbers of ordered *tentacles*, each of them touching a node. Hyperedges of hypergraphs used in this paper represent both design components and multi-argument relations among them.

By means of hypergraph transformation methods (hypergraph grammars or hypergraph operations) a large database of hypergraphs representing designs can be created. Hypergraphs can be then mapped into various graphical models of designs using a given interpretation in the form of a so called realization scheme. The interpretation specifies semantics of hypergraph elements by assigning geometrical objects and relations between them to hyperedges. The interpretation determines also transformations which design parts should undergo to be arranged properly and enables the system to create a visualization of the whole design.

Barbara Strug

Faculty of Physics, Astronomy and Applied Computer Science,
Jagiellonian University, Reymonta 4, Cracow, Poland
e-mail: barbara.strug@uj.edu.pl

Designing new artifacts requires a method of generating graphs representing them, and thus many methods for graph generation were researched. These methods include approach based on application of the theory of formal languages to the computer aided design [9, 10, 11, 23], in particular the graph based representation jointly with graph grammars [2, 9, 10, 12, 13, 22], and grammar systems [4, 6, 8, 18, 21, 25]. Other methods used to generate graphs representing designs include evolutionary computations that were used in different domains of design [2, 13, 22].

All generation methods result in producing a number of graphs representing designs. Using graph grammars and evolutionary methods, either separately or combined together, results in building a large database of graphs - or designs.

The main problem lies in the complexity and size (understood as the number of graphs) of such database. It is difficult to automatically evaluate the quality of the graphs. The quality of a graph is here understood as the quality of the design represented by this graph as a solutions to a given design problem. Thus the process of evaluation usually requires the presence of a human designer who can choose the best solution or give some numerical values to each design. The problem is that for a human "evaluator" all graphs have to be rendered to designs. This process in turn requires visualising designed objects. Unless the design problem is very simple the rendering of designs may be very complex and result in very long execution times.

A possible method of eliminating the visualization step seems to be the one using a number of graphs representing designs for which a human "evaluator" has defined a quality value as a basis for evaluating other designs in the same design problem. As it can be noticed that the designs getting higher quality values usually have some common elements, finding frequent substructures in graphs seems a useful approach.

Thus the use of methods based on frequent graph mining techniques is proposed in this paper.

2 Hypergraphs-Based Design Representation

The methods used in CAD problems like boundary representations, sweep-volume representation, surface representations or CSG (constructive solid geometry) [16, 19, 20] allow only for the "coding" of geometry of an object being designed and do not take into account the inter-related structure of many design objects i.e. the fact that parts (or components) of an object can be related to other parts in different ways. So in this paper a representations based on graphs is used. Different types of graphs have been researched and used in this domain, for example composition graphs [11, 12]. Here an extension of ordinary graphs - known as hypergraphs is used.

Hypergraphs (HGs) consist of nodes and edges, but they differ from standard graphs because the edges can connect more then two nodes, thus allowing for more complex relations to be represented. A hyperedge in a hypergraph may represent a geometrical object or a relation. A hyperedge representing an object is called object hyperedge, while the one representing a relation is called a relational hyperedge.

Nodes and hyperedges in hypergraphs can be labelled and attributed. Labels are assigned to nodes and hyperedges by means of node and hyperedge labelling functions, respectively, and attributes - by node and hyperedge attributing functions. Attributes represent properties (for example size, position, colour or material) of a component or relation represented by a given hyperedge.

3 Frequent Graph Mining Methods

Frequent graph mining techniques are developed on the basis of a more general frequent pattern mining. Frequent patterns can be defined as itemsets, subsequences, or substructures which appear in a data set with a frequency equal or greater to a user-specified threshold value. In this paper the most interesting are the substructure patterns. A substructure can refer to different structural forms, such as subgraphs, subtrees, or sublattices, which may be combined with itemsets or subsequences. If a substructure occurs frequently in a graph database, it is called a (frequent) structural pattern. Finding frequent patterns is an important part in mining associations, correlations, and many other interesting relationships among data. It may also be used in data indexing, classification, clustering, and other data mining tasks.

Frequent pattern mining was first proposed by Agrawal et al. [1] for market basket analysis in the form of association rule mining. It was used to analyze customer buying habits by finding associations between the different items that customers place in their "shopping baskets". Since the first proposal of this data mining problem and the mining algorithms relating to it, there have been many other research publications, on various kinds of extensions and applications, from scalable data mining methodologies, to handling a wide diversity of data types and a huge number of applications [30, 31].

3.1 *Apriori-Based Methods*

One of the classes of frequent subgraph mining algorithms is an Apriori method ([1]) In this approach the frequent patterns of larger size are searched for, starting from subgraphs of smaller sizes by adding to them an additional node or edge with a node.

In case of hypergraphs the size of a hypergraph is understood in this paper as the total number of hyperedges and nodes. The main reason for complexity of the Apriori-based algorithm is a candidate solution generation step. Various types of candidate solution generation algorithms were proposed. One of them is so called AGM algorithm proposed by Inokochi [17], which increases the substructure by one node at each iteration of algorithm. Two frequent graphs of size k are combined only if they have the same subgraph of size $k - 1$. The newly generated graph of size $k + 1$ includes the common subgraph of size $k - 1$ and two additional nodes from both size k graphs. As it is undetermined if there is an edge connecting the two nodes from combined graphs two different $k + 1$ size graphs are possible.

This algorithm is a basic formulation for general graphs. In case of using hypergraphs some modifications must be introduced. Firstly both the nodes and hyperedges are treated in a way only nodes are treated in standard algorithm. Moreover the tentacles (shown as "links" joining visually hyperedges and nodes) in a hypergraph are treated as edges in the algorithm.

It can be noticed that Apriori-based algorithms require a lot of computations to join two size k graphs into one size $k + 1$ graph candidate solution. To reduce these computations other methods were developed, mainly based on the pattern growth model.

3.2 *Pattern Growth Methods*

Pattern growth methodology consists in extending a single pattern into more complex one [15]. A single graph g can be extended by an edge e . This edge may or may not introduce a new node into graph. For each found frequent graph g the pattern growth algorithm tries recursively to extend it until all possible frequent graphs of which g is a subgraph are found. The recursive process ends when no more frequent graphs can be found.

The main problem of this method is the fact that each k size frequent graph can be found independently k times - by extending k different size $k - 1$ graphs.

4 Frequent Patterns in Design System

Both algorithms described in the previous section should be considered more as classes than as actual algorithms. Different algorithms were developed basing on these classes.

In this paper two algorithms, namely FFSM and gSpan, are considered. Both algorithms can work on undirected graphs with labelled nodes and edges. They perform the analysis of graphs in a depth-first order and can only find connected subgraphs. The way in which these algorithms generate potential candidates for frequent patterns is different. As each subgraph of a graph can be generated as a result of expanding different smaller subgraphs in several ways, a way of removing duplicates is also required. In a gSpan algorithm a so-called dfs-code, representing the edges in the order in which they are added to the fragment, is build. As different paths to each fragment exist also different codes are possible. By using a global order of these codes the smallest one is defined as canonical. Thus gSpan extends only those fragments which have canonical representations. To lower the cost of detecting non-canonical codes gSpan uses a heuristic based on grouping all possible extending edges into forward edges that insert a new node, and backward edges which end cycles, and then it extends the fragments with forward edges outgoing from nodes on the right most path and backward edges from the last added node. For each fragment gSpan stores a list of graphs containing a given fragment. The size of this list describes the frequency of the fragment. To extend a fragment, only the graphs from this list are used, as they contain a given fragment.

FFSM, on the other hand, represents a graph in the form of a two dimensional matrix, with node labels on a diagonal. The code is generated by concatenating all rows of the matrix. In lexicographical order isomorphic graphs have the same matrix code. Then the new fragments are generated by joining two matrices.

The approach to reasoning about designs used in this paper has been implemented and tested on examples of a floor layout design. This approach consists in several steps. In the first step a database of floor layouts was coded in the GraphML format and imported to the GraphSearcher application [29]. Then the set of frequent patterns is generated and finally these patterns are used to evaluate new designs.

The first experiment was carried out with the support set to 100%, that is for a subgraph to be considered frequent it has to appear in all hypergraphs from the database. The subsequent two experiments were carried out with the support parameter set to 90%, and 60%, respectively. In table 1 the number of frequent patterns found for various support parameters is presented.

The frequent patterns found with the support set to 100% have been subsequently used as the reasoning rules and tested against a number of new designs. Some of them are depicted in Fig. 1a and b. One of the hypergraphs representing new

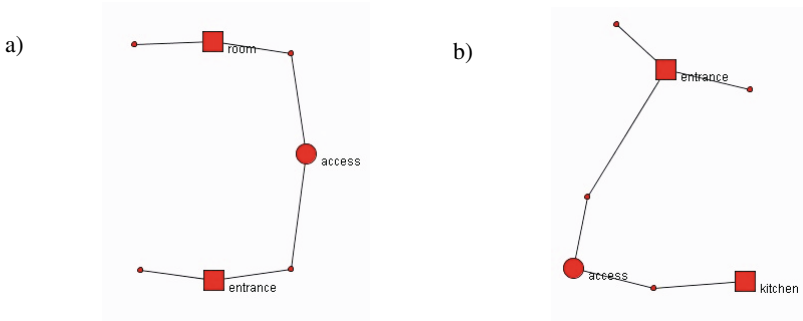


Fig. 1 Examples of frequent patterns representing design requirements

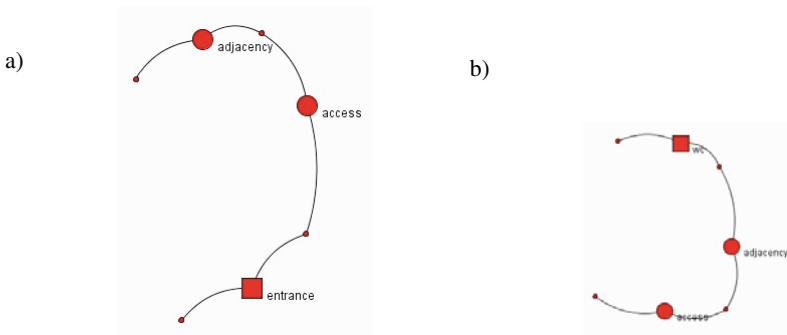


Fig. 2 Examples of frequent patterns representing meaningless rules

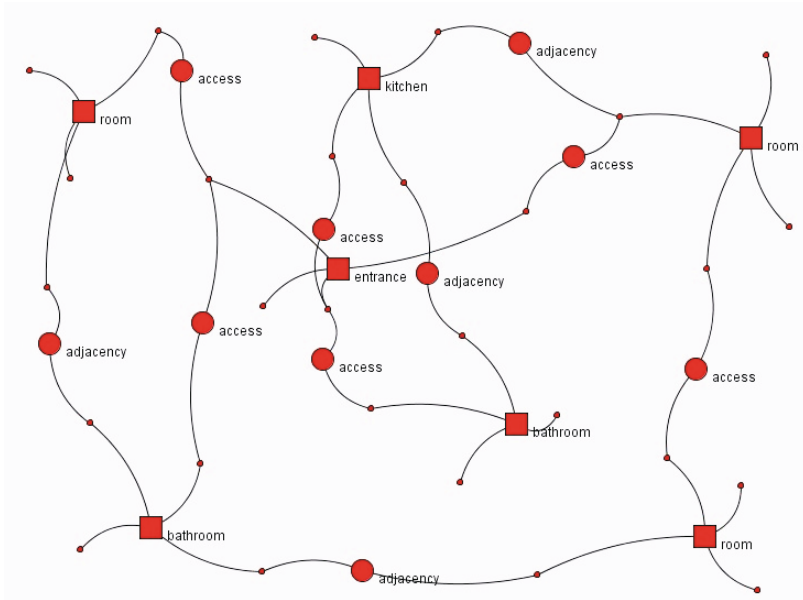


Fig. 3 Hypergraph representing a new design

Table 1 The number of frequent patterns for different values of support for FFSM and gSpan algorithms

algorithm	FFSM	FFSM	FFSM	gSpan	gSpan	gSpan
support	100%	90%	60%	100%	90%	60%
num. of freq.	910	1775	19075	1021	1808	19104

designs is presented in Fig. 3. In all figures object hyperedges are depicted as squares and relational ones - as circles. Then the designs were evaluated on the basis of the number of frequent patterns present in their hypergraph representations in such a way that those which contain the highest number of frequent substructures are considered to be the best. The quality of such a new design is expressed by a percentage of frequent patterns it contains. For a hypergraph depicted in Fig. 3 the value is 91% by FFSM-generated set and 92% by gSpan-generated set. The designs with the highest quality value are then presented to the designer for approval. Thus the automation of the rule generation process can save a lot of time in the design process.

More detailed results obtained with the use of the FFSM method were presented in [26], and the results obtained with the use of the gSpan algorithm and a comparison of both results were presented in [27].

However, two problems can be noticed. The first one is the fact that the number of frequent subgraphs even for high support parameters is very high. For the

support parameter set to 100% there were 1021 frequent patterns found by gSpan and 910 - by FFSM. As the evaluation of a hypergraph representing a new design consists in checking how many of these frequent subgraphs are also subgraphs of the new hypergraph a huge number of subgraph isomorphism checking operations is required. Even taking into account the fact that the hypergraphs are labelled, what lowers considerably the computational cost of these operations, it still is a costly and time consuming process.

The second problem is the quality of rules represented by the frequent patterns found. For example patterns depicted in Fig. 1a and b represent valid, meaningful design rules; the accessibility of a room from an entrance (Fig. 1a) and the accessibility of a kitchen from an entrance (Fig. 1b). But at the same time the frequent patterns depicted in Figs. 2a and 2b do not represent any valid design rules. Moreover the presence of many meaningless patterns affects the reliability of the evaluation - a hypergraph containing many of such patterns could get high quality value while representing an obviously bad design.

Solving the second problem i.e. eliminating meaningless frequent patterns would reduce the total number of frequent patterns and thus could at the same time contribute to solving the first problem. But it would also improve the quality and the reliability of the automatic evaluation process.

It can be observed that in all patterns that should be eliminated some common elements are present. In the frequent patterns depicted in Figs. 2a and b a presence of two consecutive relational hyperedges can be noticed. In both examples two hyperedges labelled *adjacency* and *accessibility* are connected without a single object hyperedge between them. Such a sequence of hyperedges would never represent a valid design rule. Similarly, two consecutive object hyperedges without relation between them cannot represent a correct rule. In some frequent patterns even longer sequences of same-type hyperedges were found. Thus frequent subgraphs containing such subgraphs are good candidates for elimination from the set.

Two methods of eliminating meaningless patterns were tested.

4.1 Frequent Patterns Postprocessing

First approach tested consists in analyzing all hypergraphs in the set of frequent patterns found by each algorithm and eliminating those containing unwanted subgraphs. Such a process is time consuming, what is due to a large number of graphs to analyze, but results in significant reduction of the number of patterns used in evaluation of the new designs. As this process has to be performed only once, the time it needs can be justified.

Applying this postprocessing to sets found by FFSM and gSpan algorithms with 100% support (910 and 1021, respectively) a significant reduction was achieved. In case of FFSM the reduced set consisted of 642 frequent subgraph, and for gSpan - 612. The reduction is similar for both sets, but the gSpan-generated set has been reduced slightly more. Both reduced sets have been then used to evaluate the quality of some hypergraphs representing new designs and the quality values obtained

were compared to the ones obtained with the unreduced sets. While the resulting scores are different, as was expected, the quality values are now closer to the human evaluation.

4.2 Modified Frequent Pattern Generator

In the second method the frequent pattern mining algorithm was modified by applying additional constraints to the possible extensions. It takes into account the type of the hyperedge to be added and if it is of the same type as the one to which it would be connected such an extension is forbidden and another extension is tried. As a result frequent subgraphs generated do not contain same type hyperedges sequences.

This approach, in addition to eliminating meaningless subgraphs, also speeds up the mining process. This speed-up results from smaller number of small graphs, which propagates into larger subgraphs.

The modification was so far applied only to gSpan algorithm. The set of frequent subgraphs generated by the modified algorithm consists of 585 hypergraphs. The number is slightly lower than the one obtained by the postprocessing approach, what may suggest missing some frequent patterns, but more experiments have to be carried out to verify this possibility. The reduced set has also been used to evaluate the quality of hypergraphs representing new designs and the quality values obtained were compared to the ones obtained with the unreduced set. Similarly to the first method the resulting scores are different, but the quality values are also closer to the human evaluation.

5 Concluding Remarks and Future Work

In this paper two methods of reducing sets of frequent subgraphs in hypergraphs used to represent designs are presented. These methods were tested on a set of hypergraphs representing floor layouts. The frequent patterns found were used to evaluate the quality of new designs. Both methods used resulted in reducing the size of the set of frequent patterns by approximately 30% to 40%. While it makes the evaluation of new designs much faster - as a smaller number of subgraph isomorphism tests must be carried out, the number is still large. Thus a further reduction would be desired. At the same time such a reduction has to preserve the quality of the evaluation.

One possible way of reducing the subgraph set, which is currently being investigated, is an evolutionary selection of a subset of frequent graphs. Each subset could be then evaluated by comparing the scores for a number of designs received from a full set and from a subset.

The elimination of small frequent subgraphs, i.e. subgraphs of size smaller than a predefined threshold, is also being investigated as very small subgraphs are unlikely to represent any important rules.

The modification of the FFMSM algorithm to include constraints on consecutive same type hyperedges is also planned to be investigated.

References

- [1] Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proc. 1993 ACM-SIGMOD Int. Conf. Manag. of Data SIGMOD, Washington, vol. 93, pp. 207–216 (1993)
- [2] Borkowski, A., Grabska, E., Nikodem, P., Strug, B.: Searching for Innovative Structural Layouts by Means of Graph Grammars and Evolutionary Optimization. In: Proc. 2nd Int. Structural Eng. And Constr. Conf., Rome, pp. 475–480 (2003)
- [3] Clarke, E.M., Emerson, E.A., Sistla, P.A.: Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Trans. on Progr. Lang. and Sys.* 8, 244–263 (1986)
- [4] Cshaj-Varj, E.: Grammar systems: A short survey. In: Proceedings of Grammar Systems Week 2004, Budapest, Hungary, pp. 141–157 (2004)
- [5] Cshaj-Varj, E., Vaszil, G.: On context-free parallel communicating grammar systems: Synchronization, communication, and normal forms. *Theor. Comp. Sc.* 255(1-2), 511–538 (2001)
- [6] Cshaj-Varj, E., Dassow, J., Kelemen, J., Paun, G.: Grammar systems. A grammatical approach to distribution and cooperation. *Topics in Computer Mathematics*, vol. 8. Gordon and Breach Science Publishers, Yverdon (1994)
- [7] Cshaj-Varj, E., Dassow, J., Paun, G.: Dynamically controlled cooperating/distributed grammar systems. *Information Sciences* 69(1-2), 1–25 (1993)
- [8] Dassow, J., Paun, G., Rozenberg, G.: Grammar systems. In: Salomaa, A., Rozenberg, G. (eds.) *Handbook of Formal Languages*, vol. 2(4), pp. 155–213. Springer, Heidelberg (1997)
- [9] Grabska, E.: Theoretical Concepts of Graphical Modelling. Part one: Realization of CP-graphs. *MG&V* 2(1), 3–38 (1993)
- [10] Grabska, E.: Theoretical Concepts of Graphical Modelling. Part two: CP-graph Grammars and Languages. *MG&V* 2(2), 149–178 (1993)
- [11] Grabska, E.: Graphs and designing. In: Ehrig, H., Schneider, H.-J. (eds.) *Graph Transformations in Computer Science 1993*. LNCS, vol. 776. Springer, Heidelberg (1994)
- [12] Grabska, E., Palacz, W.: Hierarchical graphs in creative design. *MG&V* 9(1/2), 115–123 (2000)
- [13] Grabska, E., Nikodem, P., Strug, B.: Evolutionary Methods and Graph Grammars in Design and Optimization of Skeletal Structures Weimar. In: Proc. of 11th ICE 2004, Weimar, pp. 145–155 (2004)
- [14] Habel, A., Kreowski, H.J.: Some structural aspects of hypergraph languages generated by hyperedge replacement. In: Brandenburg, F.J., Wirsing, M., Vidal-Naquet, G. (eds.) *STACS 1987*. LNCS, vol. 247, pp. 207–219. Springer, Heidelberg (1987)
- [15] Han, J., Pei, J., Yin, Y., Mao, R.: Mining Frequent Patterns without Candidate Generation: A Frequent-pattern Tree Approach. *Data Min. and Know. Disc.* 8(1), 53–87 (2004)
- [16] Hoffman, C.M.: *Geometric and Solid Modeling: An Introduction*. Morgan Kaufmann, San Francisco (1989)
- [17] Inokuchi, A., Washio, T., Motoda, H.: An Apriori-Based Algorithm for Mining Frequent Substructures from Graph Data. In: Zighed, D.A., Komorowski, J., Żytkow, J.M. (eds.) *PKDD 2000*. LNCS (LNAI), vol. 1910, pp. 87–92. Springer, Heidelberg (2000)
- [18] Kelemen, J.: Syntactical models of cooperating/distributed problem solving. *J. of Exp. and Theor. AI* 3(1), 1–10 (1991)

- [19] Mantyla, M.: *An Introduction To Solid Modeling*. Computer Science Press, Rockville (1988)
- [20] Martin, R.R., Stephenson, P.C.: Sweeping of Three-dimensional Objects. *CAD* 22(4), 223–234 (1990)
- [21] Martin-Vide, C., Mitrana, V.: Cooperation in contextual grammars. In: Kelemenovi, A. (ed.) *Proc. of the MFCS 1998 Satellite Workshop on Grammar Systems*, pp. 289–302. Silesian University, Opava (1998)
- [22] Nikodem, P., Strug, B.: Graph Transformations in Evolutionary Design. In: Rutkowski, L., Siekmann, J.H., Tadeusiewicz, R., Zadeh, L.A. (eds.) *ICAISC 2004. LNCS (LNAI)*, vol. 3070, pp. 456–461. Springer, Heidelberg (2004)
- [23] Rozenberg, G.: *Handbook of Graph Grammars and Computing by Graph Transformations. Foundations*, vol. 1. World Scientific, London (1997)
- [24] Rozenberg, G.: *Handbook of Graph Grammars and Computing by Graph Transformations. Applications. Languages and Tools*, vol. 2. World Scientific, London (1999)
- [25] Simeoni, M., Staniszki, M.: Cooperating graph grammar systems. In: Paun, G., Salomaa, A. (eds.) *Grammatical Models of Multi-Agent Systems*, pp. 193–217. Gordon and Breach, Amsterdam (1999)
- [26] Strug, B., Ślusarczyk, G.: Reasoning about designs through frequent patterns mining. *Adv. Eng. Inf.* 23, 361–369 (2009)
- [27] Strug, B., Ślusarczyk, G.: Frequent Pattern Mining in a Design Supporting System. *Key Eng. Mat.* 450, 1–4 (2011)
- [28] Strug, J., Deniziak, S., Sapiecha, K.: Validation of Reactive Embedded Systems against Temporal Requirements. In: *Proc. of ECBS 2004*, pp. 152–160 (2004)
- [29] Tomanek M.: Searching for graph patterns and applications. MSc thesis, Jagiellonian University (2009) (in Polish)
- [30] Yan, X., Yu, P.S., Han, J.: Graph Indexing: A Frequent Structure-based Approach. In: *Proc. of 2004 Int. Conf. on Manag. of Data SIGMOD 2004*, pp. 335–346 (2004)
- [31] Yan, X., Yu, P.S., Han, J.: Substructure Similarity Search in Graph Databases. In: *Proc. of 2005 Int. Conf. on Manag. of Data SIGMON 2005*, pp. 766–777 (2005)

A New Approach to Multi-class SVM-Based Classification Using Error Correcting Output Codes

Wiesław Chmielnicki and Katarzyna Stapor

Abstract. Protein fold classification is the prediction of protein's tertiary structure (fold) from amino acid sequence without relying on the sequence similarity. The problem how to predict protein fold from amino acid sequence is regarded as a great challenge in computational biology and bioinformatics. To deal with this problem the support vector machine (SVM) classifier was introduced. However the SVM is a binary classifier, but protein fold recognition is a multi-class problem. So the method of solving this issue was proposed based on error correcting output codes (ECOC). The key problem in this approach is how to construct the optimal ECOC codewords. There are three strategies presented in this paper based on recognition ratios obtained by binary classifiers on the training data set. The SVM classifier using the ECOC codewords constructed using these strategies was used on a real world data set. The obtained results (57.1% - 62.6%) are better than the best results published in the literature.

1 Introduction

Proteins are not just simple sequences of amino acids. They create different structures. In this paper we deal with the tertiary structure (fold) of a protein. The knowledge of it is very important because this structure of a protein is closely related to its biological function [3]. The successful completion of many genome-sequencing projects has meant that the number of proteins with known amino acids sequence is quickly increasing, but the number of proteins with known fold is still relatively

Wiesław Chmielnicki

Jagiellonian University, Faculty of Physics,

Astronomy and Applied Computer Science, Reymonta 4, 30-059 Krakow, Poland

e-mail: wieslaw.chmielnicki@uj.edu.pl

Katarzyna Stapor

Silesian University of Technology, Institute of Computer Science,

Akademicka 16, 44-100 Gliwice, Poland

e-mail: katarzyna.stapor@polsl.pl

very small. There are more than 6 millions of proteins with known amino acid sequence in the UniProt database in January 2010 and only about 58,500 proteins with known structure in the PDB database in July 2009.

Several machine-learning methods to predict the protein folds from amino acids sequences can be found in literature. Ding and Dubchak [6] experiments with Support Vector Machine (SVM) and Neural Network (NN) classifiers. Shen and Chou [18] proposed ensemble model based on nearest neighbour. A modified nearest neighbour algorithm called K-local hyperplane (HKNN) was used by Okun [16]. Nanni [15] proposed ensemble of classifiers: Fisher's linear classifier and HKNN classifier.

This paper focuses on the support vector machine (SVM) classifier. However, the SVM is a binary classifier but the protein fold recognition is a multi-class problem and how to effectively extend a binary to a multi-class classifier case is still an ongoing research problem. There are many methods proposed to deal with this issue. The most popular is to construct the multi-class classifier by combining several binary classifiers.

There are many methods proposed in literature, such as one-versus-others, one-versus-one strategies, DAG (Directed Acyclic Graph), ADAG (Adaptive Directed Acyclic Graph) methods [17], BDT (Binary Decision Tree) approach [10], DB2 method [20], pairwise coupling [11] or error-correcting output codes [7]. The last method we use in our experiments. The most important problem in this method is how to construct the optimal output code (codeword) for a class. We propose a method based on recognition ratios obtained by binary classifiers on the training data set.

The rest of this paper is organized as follows: Section 2 introduces the database and the feature vectors used in these experiments, Section 3 describes the experiments and the proposed method of constructing the ECOC matrix in Section 4 there is short discussion of the proposed method and Section 5 presents experimental results and conclusions.

2 Database

In experiments described in this paper two data sets derived from Structural Classification of Proteins (SCOP) database are used. The detailed description of these sets can be found in [6]. The training set consists of 313 protein sequences and the testing set consists of 385 protein sequences. These data sets include proteins from 27 most populated different classes (protein folds) representing all major structural classes: α , β , α/β , and $\alpha + \beta$. The training set was based on PDB_select sets ([12], [13]) where two proteins have no more than 35% of the sequence identity. The testing set was based on PDB-40D set developed by Lo Conte et al. [14] from which representatives of the same 27 largest folds are selected.

In our experiments the feature vectors developed by Ding and Dubchak [6] were used. These feature vectors are based on six parameters: Amino acids composition (C), predicted secondary structure (S), Hydrophobicity (H), Normalized

van der Waals volume (V), Polarity (P) and Polarizability (Z). Each parameter corresponds to 21 features except Amino acids composition (C), which corresponds to 20 features. The data sets including these feature vectors are available at <http://ranger.uta.edu/~chqding/protein/>. For more concrete details, see [8], [9]. In this study the sequence length was added to the Amino acids composition (C) vector and the feature vectors based on these parameters were used in different combinations creating vectors from 21D to 126D.

3 Methods

The Support Vector Machine (SVM) is a well known large margin classifier proposed by [19]. The basic concept behind the SVM classifier is to search an optimal separating hyperplane, which separates two classes. The decision function of the binary SVM is presented as:

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b \right), \quad (1)$$

where $0 \leq \alpha_i \leq C, i = 1, 2, \dots, N$ are nonnegative Lagrange multipliers, C is a cost parameter, that controls the trade off between allowing training errors and forcing rigid margins, x_i are the support vectors and $K(x_i, x)$ is the kernel function.

The LIBSVM library version 2.91 was used in our research [4]. Although the implementation of this library includes a solution for the multi category problems, only the binary version of the classifier was used. LIBSVM provides a choice of build-in kernels i.e. Linear, Polynomial, Radial Basis Function (RBF) and Gaussian. The RBF kernel:

$$K(x_i, x) = -\gamma \|x - x_i\|^2, \gamma > 0, \quad (2)$$

gave the best results in our experiments.

The parameters C and γ must be chosen to use the SVM classifier with RBF kernel. It is not known beforehand which C and γ are the best for one problem. Both values must be experimentally chosen, which was done by using a cross-validation procedure on the training data set. The best recognition ratio was achieved using parameters values $\gamma = 0.1$ and $C = 128$.

The SVM is a binary classifier but the protein fold recognition is a multi-class problem. There are many methods proposed to deal with this issue. An approach using error correcting output codes (ECOC) [7] was used in our experiments.

The ECOC technique is a general framework to handle the multi-class problems using binary classifiers. It is known that this technique can improve the generalization performance of the classifier [1], [2].

The first step to use ECOC technique is to create the output codes (codewords) for each of the N classes of the length L . Each codeword consists of sequence of 0s and 1s. In this way we build a $N \times L$ coding matrix. A column from this matrix assigns a label 0 or 1 to each class. In this way a multi-class problem is reduced to a binary

one. The second step is to train an individual binary classifier for each column of the coding matrix. These binary classifiers are used on the samples from the testing data set. After use of L binary classifiers on the sample we obtain a codeword. The sample is classified in the class with the closest codeword in the coding matrix. The closest codeword in a sense of the Hamming distance (which counts the number of bits that differ).

The main problem is how to construct the optimal codewords for the classes. We can see that every binary classifier works on the same set of samples. The only difference is a label that is assigned to each sample. So, if some of the columns are the same or similar, errors committed on this sample will be repeated many times by the classifier. In order to prevent this effect, the Hamming distance between columns should be maximized to avoid correlation between the single base classifiers. However the Hamming distance between the rows must be also maximized in order to achieve good classification ratio at the decoding phase. Due to these constraints there are four different kind of methods proposed in the literature:

- Exhaustive codes – problems with $3 \leq N \leq 7$.
- Column selection from Exhaustive codes – problems with $8 \leq N \leq 11$.
- Randomized Hill Climbing – problems with $N > 11$.
- BCH codes – problems with $N > 11$.

In our multi-class problem we have 27 classes, so first two solutions are inapplicable. The BCH codes have some practical drawbacks. For example the codes do not always exhibit good column separations and the number of rows in these codes is always a power of two. In our experiments we use a new method based on recognition ratios on the training set.

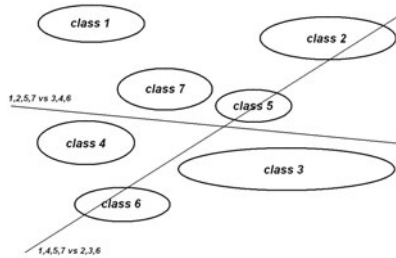


Fig. 1 Good 1,2,5,7 vs 3,4,6 and bad 1,4,5,6 vs 2,3,6 classifier

Each column of the coding matrix can be considered as a binary problem which divides two subsets of the classes. However, some of these subsets are hard to separate. The binary classifier makes a lot of errors i.e. achieve low recognition ratio. See figure 1. These columns should not be used in our coding matrix. In order to avoid this problem we divided the coding phase into two steps. First we tested randomly chosen columns on the training data set using cross-validation technique. In this way a list of the columns was created. The list was sorted according to the recognition

ratios achieved on the training data set. In our experiments we create a list of 8192 columns. The classifier related to the first column in this list achieves recognition ratio of 90.42% while the recognition ratio of the last classifier is 50.48% only.

It is seen that if we will use columns (and the binary classifiers that are related to them) that are higher in this ranking list the number of errors made by these classifiers will be relatively low. This may improve the final recognition ratio. The simplest way is to use first L columns, but these columns are usually very close considering Hamming distance and they create codewords that are very close either. The better strategy is to use only columns that are in some distance from all others. The obtained codewords are quite good in a sense of the Hamming distance, but we can try to improve this distance. We can check all columns and try to change the column used in the coding matrix by the column that precede it on the ranking list, but only when if the minimal Hamming distance will increase.

In our experiments we used four different coding matrices created by methods described below:

- Randomized Hill Climbing (RHC) - we use the algorithm described in [7].
- First Columns on the Ranking List (FCRL) - the algorithm gets first L columns from the ranking list, where L is the length of the codeword.
- Distant Columns from the Ranking List (DCRL) - the algorithm gets the first column from the ranking list. Next it searches for the column which Hamming distance from the first is at least C , then it seeks next column which Hamming distance from the first two is at least C and so on. The value of the C must be experimentally chosen, it depends of the column size (the number of classes). We choose the value of $C = 9$.
- Modified Ddistant Columns from the Rranking List (MDCRL) - the algorithm bulids coding matrix as described above and then an extra procedure is conducted. Starting from the last column the algorithm try to replace the column by the column which is higher in the ranking list, but only when if this procedure increase the minimal Hamminig distance between codewords (rows of the matrix).

In our experiments we used codewords of length of $L = 32, 64, 80, 96, 128$ and 256. The results obtained using these codewords and strategies described above are presented in the table [1].

4 Discussion

The accuracy measure used in this paper is the standard Q percentage accuracy [2]. Suppose there is $N = n_1 + n_2 + \dots + n_p$ test proteins, where n_i is the number of proteins which belongs to the class i . Suppose that c_i of proteins from n_i are correctly recognised (as belonging to the class i). So the total number of $C = c_1 + c_2 + \dots + c_p$ proteins is correctly recognized. Therefore the total accuracy is $Q = C/N$.

The crucial problem in our method is how to obtain optimal code matrix i.e. with large minimal Hamming distance between codewords. As we can see in the

Table 1 Recognition ratios obtained with different strategies and different codewords sizes

Code length	RHC	FCRL	DCRL	MDCRL
32	52.0	52.5	56.1	57.7
64	54.8	57.1	61.0	60.8
80	53.0	56.1	61.3	62.6
96	54.0	56.4	60.8	62.3
128	53.5	56.9	59.7	61.6
256	53.8	57.1	59.7	56.4

Table 2 The Hamming distances between the rows of the coding matrix obtained using different strategies and different sizes of the codewords: average - minimal

Code length	RHC	FCRL	DCRL	MDCRL
32	16.2 - 8	13.0 - 0	14.4 - 0	14.2 - 4
64	32.5 - 27	26.7 - 1	29.7 - 7	29.4 - 13
80	40.6 - 33	33.7 - 1	37.6 - 15	37.3 - 20
96	48.9 - 43	40.8 - 2	45.7 - 18	45.1 - 28
128	65.4 - 59	55.7 - 6	61.7 - 29	60.9 - 42
256	130.1 - 122	115.3 - 19	88.8 - 63	88.0 - 71

table 2 the coding matrix created using an FCRL strategy has very small values of the minimal Hamming distance between codewords. As a matter of fact, codewords of length less than 64 bits have the minimal Hamming distance equal to 0. It means that there are at least two classes that are inseparable with these codewords. However we can see that the overall recognition ratio is quite good. It is because the average Hamming distance between codewords is greater than zero, so most classes are separable as well.

Let us examine the table 2. It shows the minimal and the average Hamming distances obtained using different methods. The RHC method obtain best average and minimal Hamming distances and what is more the minimal distance is very close to the average one. The codewords are nearly optimal, but as we can see the overall results are worse than even in simple FCRL strategy which gives very poor codewords. Let us consider this experimental result. In RHC method the codewords are nearly optimal, but we use classifiers which produce very large number of the mistakes. The last classifier in our ranking list has the recognition ratio slightly more than 50% (if we use random procedure we should have 50% on the average). On the other hand in an FCRL strategy the classifiers are very good (the recognition ratios about 90%) but we use very poor codewords. For example if we use 128 bit codeword the

Table 3 Comparison among different methods

Method	Recognition ratio
SVM ([6])	56.0%
HKNN ([16])	57.4%
RS1_HKNN_K25 ([15])	60.3%
PFP Pred. ([18])	62.1%
MLP ([5])	48.8%
FCRL (this paper)	57.1%
DCRL (this paper)	61.3%
MDCRL (this paper)	62.6%

minimal Hamming distance is 6. So to distinguish between two closest codewords (and related classes) we cannot make more than $(6 - 1)/2 = 2$ mistakes on each codeword.

The DCRL strategy tries to deal with this problem. The way we construct the coding matrix produces better codewords by the cost of using less optimal classifiers. As we can see in the table 2 as a result we obtain the codewords with better minimal Hamming distances. The procedure which improves these codewords increases the minimal Hamming distance between codewords. The average Hamming distance is only slightly smaller.

5 Conclusion

In this paper we propose the new strategies to create the ECOC codewords. These strategies focus on minimizing the number of errors made by individual binary classifiers instead of minimizing the minimal Hamming distance between codewords. We show that this strategy gives better results than simple RHC strategy.

The results achieved using the proposed strategies are promising. The recognition rates obtained using these algorithms (57,1% - 62,6%) are comparable or better than those described in the literature (48,8% - 62,1%).

The obtained results are very encouraging. Our results improved the recognition ratio achieved by other methods proposed in the literature but however some extra experiments are needed especially to consider other strategies to improve minimal Hamming distance between codewords.

References

- [1] Allwein, E., Schapire, R., Singer, Y.: Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research* 1, 113–141 (2002)
- [2] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C., Nielsen, H.: Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 16, 412–424 (2000)

- [3] Chan, H.S., Dill, K.: The protein folding problem. *Physics Today*, 24–32 (February 1993)
- [4] Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001), <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [5] Chung, I.F., Huang, C.D., Shen, Y.H., Lin, C.T.: Recognition of structure classification of protein folding by NN and SVM hierarchical learning architecture. In: Kaynak, O., Alpaydin, E., Oja, E., Xu, L. (eds.) ICANN 2003 and ICONIP 2003. LNCS, vol. 2714, pp. 1159–1167. Springer, Heidelberg (2003)
- [6] Ding, C.H., Dubchak, I.: Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics* 17, 349–358 (2001)
- [7] Dietterich, T.G., Bakiri, G.: Solving multiclass problems via error-correcting output codes. *Journal of Artificial Intelligence Research* 2, 263–286 (1995)
- [8] Dubchak, I., Muchnik, I., Holbrook, S.R., Kim, S.H.: Prediction of protein folding class using global description of amino acid sequence. *Proc. Natl. Acad. Sci. USA* 92, 8700–8704 (1995)
- [9] Dubchak, I., Muchnik, I., Kim, S.H.: Protein folding class predictor for SCOP: approach based on global descriptors. In: *Proceedings ISMB*, vol. 5, pp. 104–107 (1997)
- [10] Fei, B., Liu, J.: Binary Tree of SVM: A New Fast Multiclass Training and Classification Algorithm. *IEEE Transaction on Neural Networks* 17(3), 696–704 (2006)
- [11] Hastie, T., Tibshirani, R.: Classification by pairwise coupling. *Annals of Statistics* 26(2), 451–471 (1998)
- [12] Hobohm, U., Scharf, M., Schneider, R., Sander, C.: Selection of a representative set of structures from the Brookhaven Protein Bank. *Protein Sci.* 1, 409–417 (1992)
- [13] Hobohm, U., Sander, C.: Enlarged representative set of Proteins. *Protein Sci.* 3, 522–524 (1994)
- [14] Lo Conte, L., Ailey, B., Hubbard, T.J.P., Brenner, S.E., Murzin, A.G., Chotchia, C.: SCOP: a structural classification of protein database. *Nucleic Acids Res.* 28, 257–259 (2000)
- [15] Nanni, L.: A novel ensemble of classifiers for protein fold recognition. *Neurocomputing* 69, 2434–2437 (2006)
- [16] Okun, O.: Protein fold recognition with k-local hyperplane distance nearest neighbor algorithm. In: *Proceedings of the Second European Workshop on Data Mining and Text Mining in Bioinformatics*, Pisa, Italy, September 24, pp. 51–57 (2004)
- [17] Platt, J.C., Cristianini, N., Shawe-Taylor, J.: Large Margin DAGs for Multiclass Classification. In: *Proceedings of Neural Information Processing Systems*, pp. 547–553 (2000)
- [18] Shen, H.B., Chou, K.C.: Ensemble classifier for protein fold pattern recognition. *Bioinformatics* 22, 1717–1722 (2006)
- [19] Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
- [20] Vural, V., Dy, J.G.: A hierarchical method for multi-class support vector machines. In: *Proceedings of the Twenty-First ICML*, pp. 831–838 (2004)
- [21] Windeatt, T., Ghaderi, R.: Coding and decoding for multiclass learning problems. *Information Fusion* 4(1), 11–21 (2003)

Fractal Dimension in the Landscape Change Estimation

Piotr Łabędź and Agnieszka Ozimek

Abstract. In recent years, fractal dimension has been applied to medical images analysis, material engineering, as well as to satellite photography of the terrain. In the paper the attempt is made to use fractal dimension as an objective indicator supporting the assessment of landscape changes predicted as a result of new anthropologic elements introduction. Authentic expert assessments have been analysed and comparisons have been made between the intuitive evaluations and results of calculations.

1 Introduction

Landscape is a common good, therefore it has to be protected, particularly in the most attractive parts. It should be mentioned here that about 30% of the terrain in Poland is under legal preservation. In these areas every change, e.g. new building creation, may lead to their aesthetic degradation. Thus, the investigations are usually conducted, in order to evaluate the most beautiful views and to assess the extent of impact that the potential investment would have on them.

In this process, different types of approaches have been represented. Generally they can be divided into three groups: expert opinions (professional, formal, aesthetic), perception-based (empirical, phenomenological, psychological) and cognitive ones (behavioural, psycho-physical, quantitative holistic models) [8] provide a comprehensive description of these methods. In Poland, the first type of attitude is commonly represented. Every resolution concerning a new investment, located in the context of the protected landscape, requires a great deal of intimate and complex knowledge from an expert, without any guarantee of objective results, because the assessment bases primarily on his intuition [9]. Though landscape perception

Piotr Łabędź · Agnieszka Ozimek
Institute of Computer Science
Cracow University of Technology
ul. Warszawska 24, 31-155 Cracow, Poland
e-mail: [{{plabedz, aozimek}@pk.edu.pl](mailto:{plabedz, aozimek}@pk.edu.pl)

is a difficult object for parameterisation, objective tools supporting this process of decision making are essential [1].

Recently, a number of ideas arose, concerning possibilities of fractal dimension application in image analysis. These attempts were made for medical imaging (retina, MRI) [6], satellite photographs of the terrain [3] or industrial parts and materials [4]. Therefore, fractal dimension seems to be potentially useful in views evaluation. While the expert has to judge the impact that a new element has on the existing environment, comparison between the current view and the future one should be made. The numerical values obtained as the difference of fractal dimensions, calculated for two photographs ("before" and "after") or among different variants of the investment, would be helpful in decision making or choosing the best version of project.

2 Research Method

2.1 Box-counting Dimension

Dimension is a term that is not easy to define. Generally, dimension of an object (called topological dimension) is understood as a number of independent parameters (coordinates) that are necessary in order to determinate its vertices unambiguously. In fractal geometry, it is more adequate to use other definitions, e.g. Hausdorff dimension, self-similarity, compass dimension, box-counting. Sometimes for their measurement, additional conditions have to be met: e.g. for the self-similarity dimension it would be the mathematical or geometrical self-similarity existence, whereas for the compass dimension the assumption is that the measured object is in the form of a curve [2]. For landscape photographs analyses these conditions are almost impossible to fulfil, so the box-counting dimension seems to be the most useful. Moreover, two-dimensional digital photographs are represented as a matrices - the type of data on which calculations with the use of two-dimensional "boxes" are relatively easy to perform.

In order to calculate the box-counting dimension, we should locate a structure on regular grid with the mesh dimension s and count these meshes that contain even the smallest part of the structure. Thus we receive the number of filled meshes N which is connected with their dimension, for that reason it is registered as $N(s)$. Subsequently, we diminish size of the meshes and repeat the operation. In practice, the algorithm can be continued until we are restricted by the device resolution (Fig. 1).

After the whole procedure we could plot obtained data. Horizontal axis scale would be logarithms of the mesh dimension inverses and vertical axis - logarithms of filled meshes numbers (Fig. 2). It appears that the straight line of regression can be fit comparatively easy to this set of data [5].

Box-counting dimension. D_b is defined as a coefficient of regression line inclination on the graph of relationship between the number of filled boxes (meshes) and their size [7].

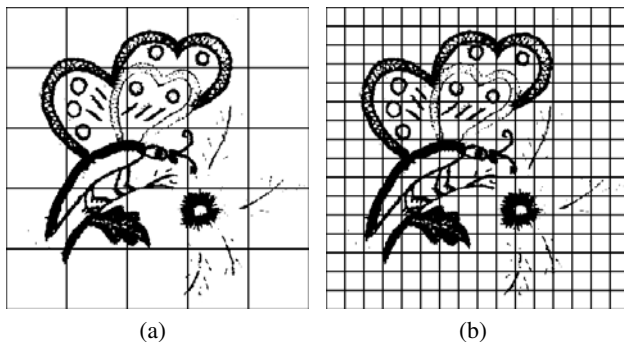


Fig. 1 Structure covered with a mesh dimension (a) $s = \frac{1}{5}$ of the whole, and (b) $s = \frac{1}{16}$

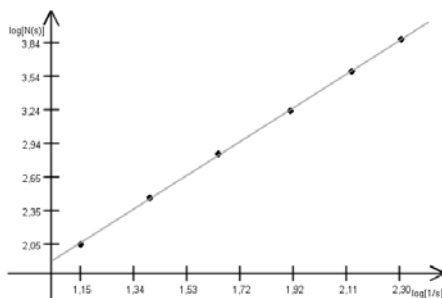


Fig. 2 Dependence between number of filled meshes and their dimension

Formal definition

The definition presented above is quite easy to practical implementation, whereas more strict definition (regarding to mathematical formalism) was presented by Falconer.

The box-counting dimension of a subset F of \mathbb{R}^n is given by

$$D_b(F) = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{\log \frac{1}{\delta}} \tag{1}$$

(if this limit exists), where $N_\delta(F)$ is any of the following:

- the smallest number of closed balls of radius δ that cover F ;
- the smallest number of cubes of side δ that cover F ;
- the number of δ -mesh cubes that intersect F ;
- the smallest number of sets of diameter at most δ that cover F ;
- the largest number of disjoint balls of radius δ with centres in F .

Thereabout equation (1) says that $N_\delta(F) \simeq \delta^{-s}$ for small δ , where $s = D_b(F)$. More precisely

$$N_\delta(F)\delta^s \rightarrow \begin{cases} \infty & \text{if } s < D_b(F) \\ 0 & \text{if } s > D_b(F) \end{cases} \quad (2)$$

But

$$N_\delta(F)\delta^s = \inf \left\{ \sum_i \delta^s : \{U_i\} \text{ is a (finite) } \delta\text{-cover of } F \right\} \quad (3)$$

which could be compared with the definition of Hausdorff dimension and measure. Therefore box-counting dimension could be treated as some kind of simplification of Hausdorff dimension, which tend to be easier to calculate [2].

2.2 Implementation

Every application designed for box-counting dimension calculation, that authors tested, operated only on black-white images. In order to make measurement on colour images possible, the new plug-in for GIMP graphics program was developed. It analyses picture pixel by pixel in order to state whether single pixel belongs to an object or to the background. Afterwards, calculations are conducted according to the algorithm described above. With regard to raster construction of digital image it seems reasonable to reverse the procedure - instead of diminishing box sizes it is better to start from the smallest achievable part i.e. single pixel. In the successive steps box size is increased according to the values set in the program. After achieving preset number of steps regression line is calculated, using the method of the smallest squares. The results obtained are within a range of 1.0 to 2.0. Although the created plugin enables image segmentation to discern an object from a background, which is based on colour range estimation, the experiments revealed that manual segmentation is more precise.

3 Research and Results

3.1 Preliminary Researches

Fractal dimension, on account of its mathematical objectivity, could be used in various analyses similar to the ones presented in this paper. It should be noticed that there is no authors' intention to calculate exact value of box-counting dimension forasmuch it is hard to procure from discrete raster image. Depending on preset plug-in values, slight differences in the final results can appear, which are related to many factors. Another reason is that fractal dimension of two complete different objects could be approximate as shown in Fig. 3.

In order to verify usability of box-counting dimension in estimating changes performed in landscape and registered on photographs, various tests on simple figures were made at first. Let us consider a black rectangle on a white background as in



Fig. 3 Calculated box-counting dimension of both objects equals 1,622

Fig. 4(a), which measured box-counting dimension equals 1,903 (it would be precisely 2,0 if there was no background, so the rectangle filled entire available space). If we add another rectangle to a picture (Fig. 4(b)), which is similar in character, alteration of measured dimension would be insignificant (0,002).

Otherwise, an object that is disparate (e.g. straight line - Fig. 4(c)) induced much more reduction of calculated value.

After disparate objects multiplication, we obtain a picture that is different in character and therefore its box-counting dimension decreases at almost 25 percent (Fig. 4(d)). Moreover, we could take into consideration a picture composed only of several straight lines as in Fig. 4(e). Appending figures congruent in character (Fig. 4(f)) affects the calculated value slightly, while adding a different object, such as a rectangle, has a great influence on it.

3.2 *Landscape Photograph Analysis*

After tests that were made on artificial images, analyses on real landscape pictures were performed. One of case studies was related to the hotel in the old city centre vicinity. On account of unmendable constructional errors the decision was made that it should be demolished.

Regarding highly attractive location of the building plot, the developer desired utilizing it to the maximum possible extent. Three different proposals of arrangement were presented and subjected to evaluation with the use of box-counting dimension. Each of them is produced on Fig. 5 as a binary image containing only buildings that are marked black.

As a point of reference, the existing view (Fig. 5(a)) was taken with calculated dimension 1,664. It is worth to notice that current structure of the hotel does not harmonize very well with the surroundings, however after almost forty years it was generally accepted. The first project version assumed one long edifice that covered entire building plot with the height alike existing hotel (Fig. 5(b)). Fractal dimension rises for that proposal to 1,705. The high structures dominating the surrounding area (Fig. 5(c)) implicated the increase of the calculated value to 1,738. For the last

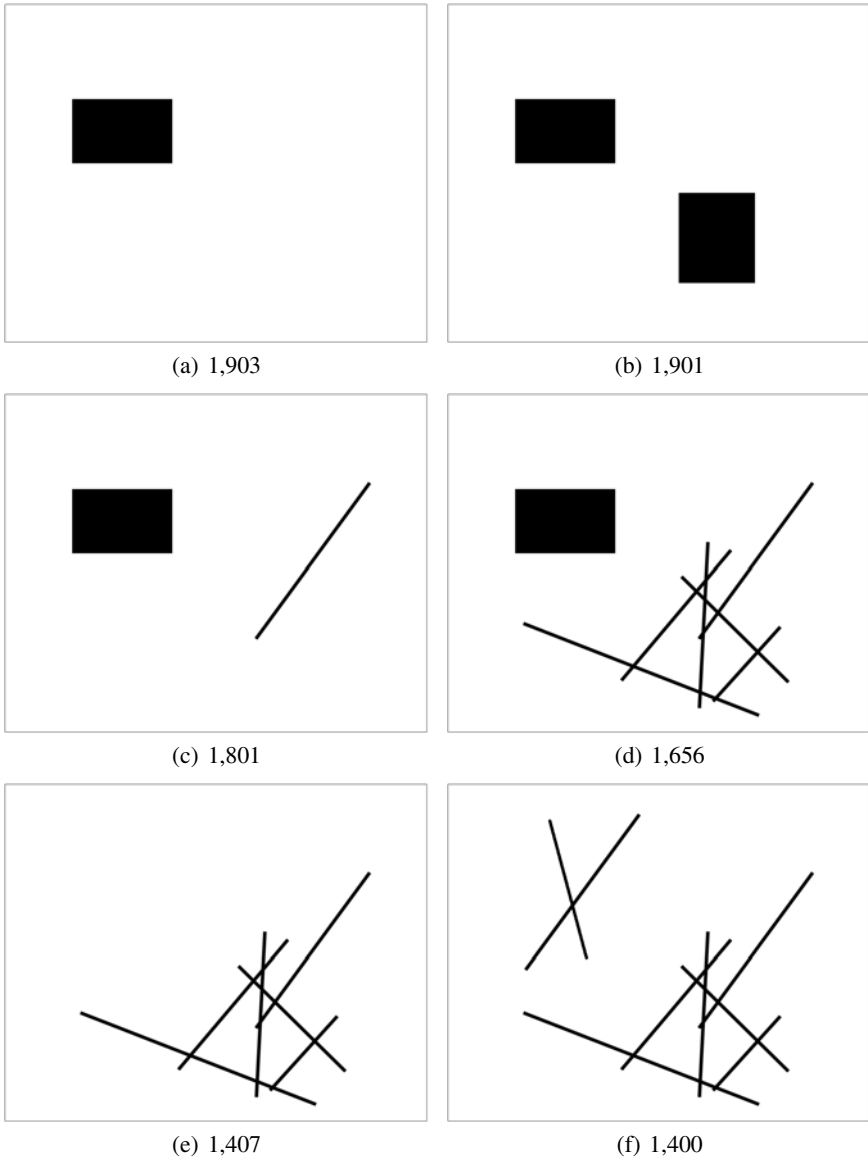
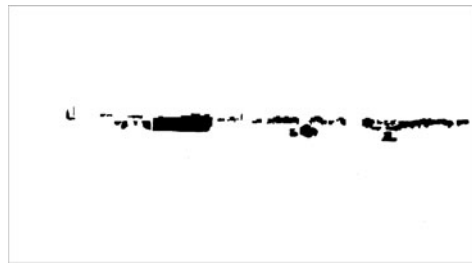
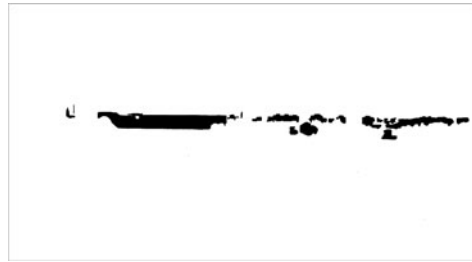


Fig. 4 Figures and their fractal dimensions

variant of the project (Fig. [5\(d\)](#)), with structures similar in dimensions to the surrounding buildings, box-counting dimension equals only 1,647. The smallest difference, as well as the reduction of the calculated value, indicate that the change in landscape (perceived from this vantage point) would be the slightest and proposed structures harmonize the best with the surroundings.



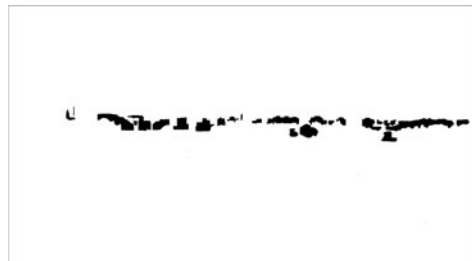
(a)



(b)



(c)



(d)

Fig. 5 Different proposals of arrangement of the area on the bank of the Wisła River with buildings marked black: (a) present view - the Forum Hotel visible as the biggest object; (b) proposal #1 - one long edifice of a height similar to the existing hotel; (c) proposal #2 - very high structures dominating the city panorama; (d) proposal #3 - several buildings that have dimensions similar to the surroundings



Fig. 6 View of the Forum Hotel

The comparison between "current" view and three suggested variants has confirmed intuitive opinions. Analogical results were obtained in various landscape photographs analyses performed by authors and not presented in this paper, including undermentioned:

- the footbridge that will connect the banks of the Wisła River
- two different localisations of wind farms in the South-East Poland

4 Conclusions

In this paper, an effort toward using box-counting dimension in landscape photographs analyses has been presented. Tests performed on artificial images indicated a relationship between changes introduced to an image and its fractal dimension. Moreover, the tests suggested that the scale of change depended on the character of modification. Preliminary results were confirmed by analyses performed on landscape photographs, regarding difference of fractal dimensions, calculated for two photographs ("before" and "after") as well as the character of change. All cases were the subjects of expert opinions in which they gave the significant criterion, supporting decision making, because of their mathematical objectiveness. It should be noticed that the core idea of the method lies in the comparison between images, therefore calculation of the exact value of box-counting dimension of objects visible in the photograph is not authors' intention. Presented method cannot constitute the only foundation for expert opinions, as well.

References

- [1] Böhm, A.: Planowanie przestrzenne dla architektów krajobrazu: Oczynniku kompozycji. Wydawnictwo Politechniki Krakowskiej, Kraków (2006)
- [2] Falconer, K.: Fractal Geometry: Mathematical Foundations and Application. John Wiley & Sons, Chichester (2003)

- [3] Ibáñez, J.J., Pérez-Gómez, R., San José Martínez, F.: The spatial distribution of soils across Europe: A fractal approach. *Ecological Complexity* 6, 294–301 (2009)
- [4] Jahn, R., Truckenbrodt, H.: A simple fractal analysis method of the surface roughness. *Journal of Materials Processing Technology* 145, 40–45 (2004)
- [5] Kudrewicz, J.: *Fraktale i chaos*. WNT, Warszawa (1993)
- [6] Oczeretko, E.: *Wymiar fraktalny w analizie sygnałów i obrazów biomedycznych*. Wydawnictwo Uniwersytetu w Białymstoku, Białystok (2006)
- [7] Peitgen, H.O., Jürgens, H., Saupe, D.: *Granice chaosu: fraktale (t. 1 - 2)*. PWN, Warszawa (2002)
- [8] Smdardn, R.C., Palmer, J.F., Felleman, J.P.: *Foundations for Visual Project Analysis*, New York (1986)
- [9] Wejchert, K.: *Przestrzeń wokół nas*. Fibak Noma Press, Katowice (1993)

An Oversegmentation Method for Handwritten Character Segmentation

Magdalena Brodowska

Abstract. This paper presents a method for generation of set of potential splitting paths for a word image. It combines some existing oversegmentation techniques with novel approaches based on joint use of word profile extrema and external background analysis and on word image skeleton. It can be used as initial segmentation step in most systems that are able to take advantage of oversegmentation. Experiment conducted for this article shows that this method is able to achieve reasonable compromise between oversegmentation rate and accuracy measured as percentage of correctly detected segmentation points.

1 Introduction

Off-line recognition of cursive handwriting can be employed in a variety of real word tasks. Many of them (e.g. reading addresses from envelopes, analyzing bank cheques) requires very high level of reliability. This and the substantial difficulty of the task gives explanation to the broad range of different methods performing cursive word recognition. Character segmentation (i.e. cutting word image into fragments corresponding to single characters) is an integral part of most of them. Importance of this steps comes from the fact, that it is exceedingly difficult to recognize improperly extracted character. On the other hand, performing segmentation correctly without classification feedback is hardly possible (e.g. discerning handwritten "rr" from n without context). This is why recognition-based segmentation approaches predominate in recent years. Most of them employ the oversegmentation approach - some set of potential cuts is generated first (with cardinality preferably greater than number of expected letters in analyzed word) from which segmentation hypotheses are obtained and evaluated with the help of classifier. General schema of such system is provided in Fig. [1](#).

Magdalena Brodowska

Faculty of Physics, Astronomy and Applied Computer Science,
Jagiellonian University, Reymonta 4, Krakow, 30-059 Poland
e-mail: m.dudek@uj.edu.pl

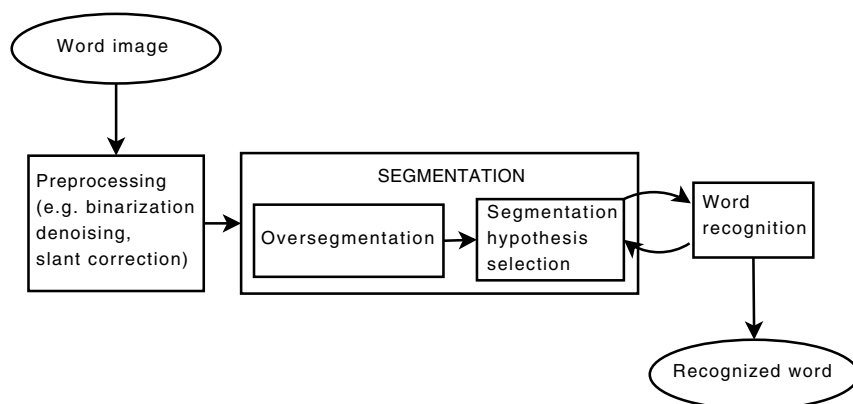


Fig. 1 General model of OCR system based on oversegmentation

There are already many distinct methods for finding segmentation paths or points described in the literature. Earliest and simplest of them were based on projection analysis (i.e. analysis of vertical or angled histogram: [11], [15], [3], [13]). They were sufficient for machine-printed words and handwriting restricted to block letters but numerous shortcomings made them inefficient in case of cursive handwriting. However, due to simplicity and low computational complexity, they are still used for preliminary and redundant generation of split points. Currently, approaches utilizing profile and contour of word image are the most common type [1, 6, 7, 8]. They are usually based on "drop falling" or "min-max" algorithms. First one places split points in local minima of profile, second one - in minima of upper profile or contour and maxima of lower profile or contour. Oliveira proposed in [9] an interesting method of numeral string segmentation in which contour analysis is supported by the skeleton analysis. The algorithms examining skeleton of the image background [4, 5, 14] or so called background regions [12, 15] are most common examples of another group - methods based on background analysis. Those methods originate from the observation, that correct segmentation path should be led between characters, that is - in background, only occasionally intersecting written lines.

This paper presents an oversegmentation method that can be applied in recognition systems dedicated to unconstrained cursive handwriting. Chapter 2 contains some definitions used in the paper. Chapter 3 describes the way of finding candidate segmentation paths based on word profile and skeleton analysis. Experimental results are presented in chapter 4.

2 Basic Terms and Definitions

Word profile

Word profile is an image obtained by retaining only outermost foreground pixels in each column of binarized image. If such achieved line is not continuous, the gaps are

filled with vertical lines. The shape made of pixels closest to upper edge of image is called *upper profile*, while the one made of pixels closest to bottom edge is called *lower profile*.

Upper background

Intuitively, upper background is the part of the binarized image background placed above foreground lines. It is calculated as follows. Empty (i.e. containing only background) columns are removed from both sides of image. Next a single background row is added at the top of the image. Upper background is a set of background pixels accessible from the added row without crossing foreground pixels.

Lower background

Analogically, lower background is the part of the binarized image background placed below foreground lines. It is calculated identically to upper background, but the background row is added at the bottom of image instead at the top.

External background

External background is a set of pixels belonging either to upper or lower background. It contains all of the background pixels with exception of those placed inside loops. Notice that upper and lower backgrounds are separate only if they are calculated for image with exactly one connected component.

Average stroke width W

Average stroke width is estimated from two quantities taken from binarized image: number of foreground pixels N_f and number of edge pixels N_e (i.e. foreground pixels with at least one background neighbor):

$$W = 2 \cdot N_f / N_e \quad (1)$$

Explanation and derivation of this formula can be found in [10].

3 Generation of Candidate Segmentation Paths

Finding potential segmentation paths in method proposed in this paper is based on profile and skeleton of the word and its background. Paths created by this solution are divided among four categories. First one represents natural gaps between letters, the others correspond to three types of connections between letters. Those categories are described below. The input for this step is the binarized image of the word.

3.1 Paths between Connected Components

In handwritten word, there are often gaps between neighboring characters, which makes segmentation easier. On the other hand, some letters consist of several disconnected components (usually dots of "i" and "j" or line with a gap). Therefore, instead of detecting fragmented letters, paths between connected components are treated just like any other segmentation paths - segmentation hypotheses. Since boxes of distinct adjacent components might overlap, straight lines cannot be used as paths between them. Instead, they are drawn along lines of skeleton of the image background. Background skeleton is generated with method described in [2] with reversed colors (see Fig. 2). Then it is adjusted by removing all "loose fragments" (i.e. all pixels with at most one foreground neighbor are removed recursively till there is no more such pixels in the skeleton). In the next step, paths that go from the uppermost line of the skeleton to the lowermost line are obtained. They form a set of segmentation paths of the first category. Lines nearest to the left and right edges of the word image are not included in the set since there are no components between them and the image bounds.

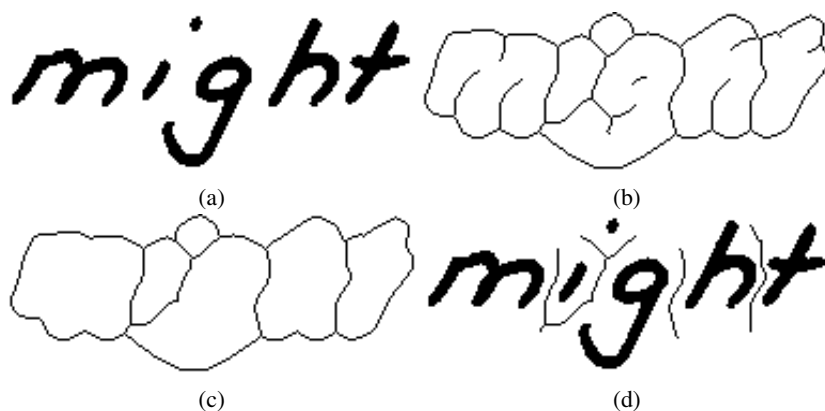


Fig. 2 (a) word image; (b) skeleton of the image background; (c) skeleton with removed "loose fragments"; (d) segmentation paths between connected components

3.2 Paths between Pairs of Profile Extrema

Segmentation paths belonging to the second category are determined by analyzing profiles of the word. First, certain characteristic points of the profiles are found. In case of upper profile, those characteristic points are simply local minima (denoted later UPM - upper profile minima). For lower profile the characteristic points include both minima and maxima (denoted LPE - lower profile extrema).

Next for each UPM a check is performed: if under such point, in a distance close to average stroke width (W) an LPE is found, then a segmentation path is drawn

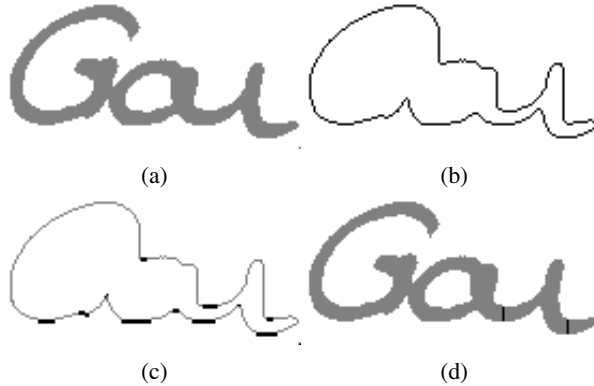


Fig. 3 (a) word image; (b) word profile; (c) UPM and LPE points (black); (d) segmentation paths between pairs of profile extrema

between those two points. If many neighboring pixels are part of single local extremum (plateau), then whole such group is treated as one characteristic point. See Fig. 3 for example of path of this category.

Determining the paths in such way aims at separating of characters connected in one point or by lower convex ligatures. Such connections are common in handwriting based on Latin alphabet. Because coupling is made between two profiles, we can avoid analyzing some of the places that, although having profile extrema, should not be segmentation points (e.g. inside loops, on the lower ends of vertical strokes). Of course, paths made according to such rules will intersect the valleys of letters u, w, v and y. This, however, is compliant with oversegmentation idea, as it delays the decision on difficult cases (like discerning single 'u' from digraph 'ii') to the point when additional information from the classification step is available.

3.3 Paths between Profile Extrema and External Points

Method of finding segmentation paths described above is sensitive to slant. Of course, it is possible to apply slant correction before segmentation step, but it may not be sufficient, especially if there is not enough handwriting available from one author to measure angle of slant reliably. Moreover, this angle does not always stay the same within one document, or even word. Pair of corresponding characteristic points may also be omitted due to covering one of them by overlapping strokes of neighboring letters in case of straight but sprawling handwriting. To deal with these problems segmentation paths of the third category are created in places where yet unused profile extrema lie near the opposite contour of word (preferably the distance is near W). UPM points that do not belong to any pair created in point 3.1 are analyzed first. For each of those points, a triangle with base oriented horizontally and placed below UPM, the vertex opposite to base being placed in UPM in question is created. The size of the triangle is calculated from average stroke width W (in

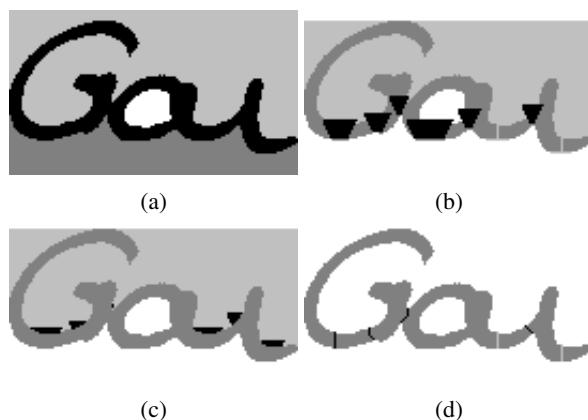


Fig. 4 (a) upper background (light gray) and lower background (dark gray); (b)-(c) construction of triangles for LPEs; (d) segmentation paths of category 2 (light gray) and 3 (black)

experiments isosceles triangles with altitude of $1.5W$ and base length of $2W$ were used). If there is at least one point of bottom background inside the triangle, a candidate segmentation path is drawn through that point and the analyzed UPM. If there are more than one bottom background points inside the triangle, one with minimum distance to UPM is chosen. LPE point are analyzed analogically, but the triangles have horizontal bases above the third vertex (placed in LPE), and top background points are sought inside them. For example, see Fig. 4.

Candidate segmentation paths created in this way allow, apart from detecting connections omitted in previous category, to separate characters connected with oblique or nearly horizontal ligatures, which commonly have profile extrema at their ends.

Both categories of profile based paths guarantee separation of single connected components into two disjoint ones (which is a trivial property derived from definitions of lower and upper profiles and top and bottom backgrounds). The restrictions on distance between points prevent accidental cutting of other lines.

3.4 Paths between Skeleton Cross Points

Segmentation paths described above are able to separate characters that touch each other in one point or are connected with an additional line (ligature). Such cases constitute majority of desired segmentation points, however it is not entirely uncommon for adjacent characters to touch in more than one place or have common line fragment. In such situations we have to deal with crossing lines, so using word skeleton for their detection is a natural choice.

First, a skeleton of the word is found with method described in [2]. Next, it undergoes simple morphological processing by applying masks from Fig. 5a. Each mask is applied simultaneously (only to every foreground pixel of the skeleton, the pixel being in the middle - to save resources). If a mask fits the analyzed

fragment, the pixel in the middle is changed to background. This step is repeated for a number of times dependent on average stroke width (namely $2W$ in tests). This operation, which can be seen as a specialized form of erosion, shortens the "loose ends" of skeleton by one pixel in each repetition (see Fig. 6). Short fragments of skeleton, such as queues or thinning algorithm artifacts appearing due to variable line width, disappear completely along with their crossing points with main skeleton. Such crossing points are never a beginning of new letter, so they are spurious with respect to character segmentation and need not be analyzed. Alternatively, thinning could be preceded by smoothing of the image with closing and opening operations. Such method, however, smoothes only very small disturbances, usually much thinner than average stroke width, and yet changes input image in a way that has potential for generating errors at thinning step (joining of lines that were separate etc.). Another advantage of the proposed transformation is that it needs only to be applied to foreground skeleton pixels, instead of the whole image, so is more computationally efficient.

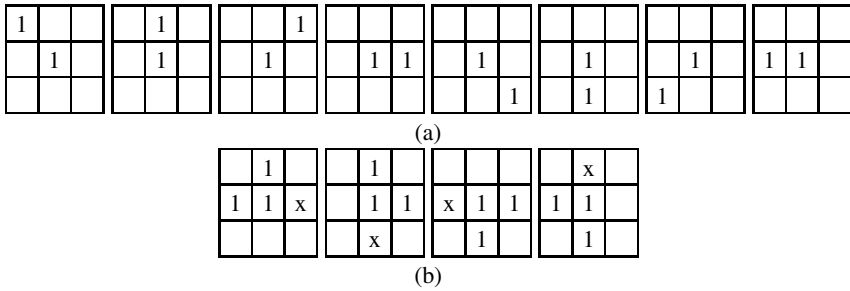


Fig. 5 Skeleton masks (1 denotes foreground pixels, empty box - background pixels, x - color is not important)

As a next preparation step, word skeleton undergoes transformations by masks shown in Fig. 5b. Again, each mask is applied simultaneously for every foreground pixel of skeleton (with this pixel being in the center of the mask) and if the mask fits - the pixels becomes background. This operation is done only once. Its purpose is to reduce the set of pixels with at least three foreground neighbors only to the places where actual crossing of lines occur, eliminating the places that are only orthogonal change of line direction (see Fig. 6b). In such prepared skeleton, all points with at least 3 neighbors are found (denoted later as SCP - skeleton cross point). For each SCP algorithm checks if there are any other SCPs inside a rectangle with width of $3W$ and its vertical symmetry axis going through the SCP. If it happen so, such points make a group of SCPs (which denoted GSCP), and they will be elements of a new segmentation path. SCPs from GSCP are sorted by their vertical coordinate. The top pixel becomes linked with closest pixel of upper background, the bottom one - with the closest point of lower background. The part of the path in between those two points runs from one SCP from that GSCP to another. The exact way of drawing this line is quite complicated and it depends on the type of

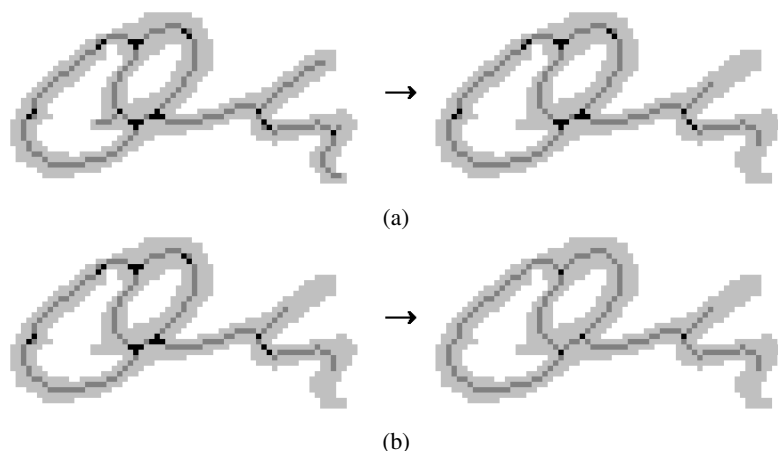


Fig. 6 Skeleton (dark gray) and its cross points (black) after applying masks operations

connection. If adjacent characters have common line fragment (which happens often when neighboring loops touch), the cut should run along the skeleton. In other cases, separating along the contour is a better idea. It is possible that both cases occur in one GSCP (see Fig. 7).

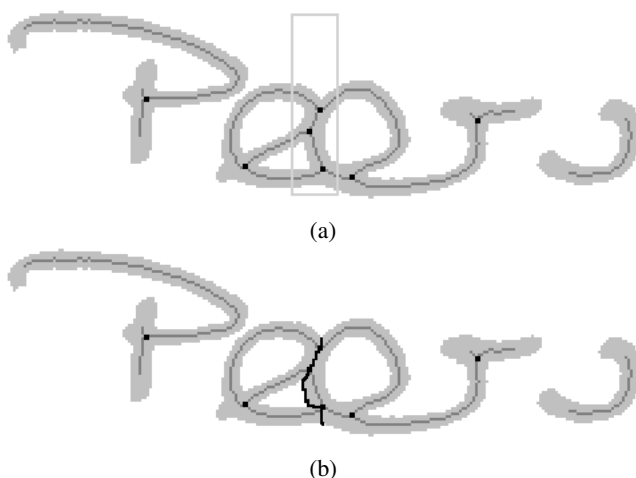


Fig. 7 (a) skeleton cross points and GSCP; (b) segmentation path of category 4

Automatic decision, which case we are dealing with is somewhat problematic. In proposed method, a following decision algorithm was applied. For a given SCP, a center of weight of a figure made from all branches of skeleton going out of

this point with length $2W$ is calculated. If it lies above the SCP, the fragment of segmentation paths leading to next SCP is drawn along the skeleton, otherwise, it is drawn along the contour lying on the left side of both SCPs (example of segmentation path can be seen in Fig. 7).

4 Experimental Results

Presented method was tested on a set of 500 words (2146 characters) obtained randomly from the NIST database. The results are in Table 1.

Table 1 Oversegmentation results

Correctly segmented words (%)	96.00
Bad segmentation (%)	1.58
Oversegmentation ratio	1.99

We consider a word correctly segmented if all correct segmentation paths are included in the result set. As this is an oversegmentation generation method, and not a final segmentation, splitting single character into multiple segments is not considered as error. Oversegmentation ratio equals ratio of total number of proposed segments to the total number of letters in testing set. Bad segmentation label includes both missing segmentation point case and a case when misplaced segmentation path makes character recognition very hard. The error ratio is smaller than those reported in [1], [13] and [8]. This, however, comes at a cost of somewhat increased number of segmentation hypotheses, especially when compared to [1] and [13]. Oversegmentation ratio is close to the one of method [8].

References

- [1] Bozekova, M.: Comparison of Handwritings. Diploma thesis, Comenius University, Bratislava, Slovak Republic (2008)
- [2] Huang, L., Wan, G., Liu, C.: An Improved Parallel Thinning Algorithm. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition (2003)
- [3] Leedham, C.G., Friday, P.D.: Isolating individual handwritten characters. In: Proc. IEE Colloq. Character Recognition and Applications, London (1989)
- [4] Liang, Z., Shi, P.: A metasynthetic approach for segmenting handwritten Chinese character strings. *Pattern Recognition Letters* 26, 1498–1511 (2005)
- [5] Lu, Z., Chi, Z., Siu, W., Shi, P.: A background-thinning-based approach for separating and recognizing connected handwritten digit strings. *Pattern Recognition* 32, 921–933 (1999)
- [6] Madhvanath, S., Kim, G., Govindaraju, V.: Chaincode Contour Processing for Handwritten Word Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(9), 928–932 (1999)

- [7] Morita, M., Lethelier, E., Yacoubi, A.E., Bortolozzi, F., Sabourin, R.: An HMM-based Approach for Date Recognition. In: Proceedings of the Fourth IAPR International Workshop on Document Analysis Systems (2000)
- [8] Nicchiotti, G., Scagliola, C., Rimassa, S.: Simple And Effective Cursive Word Segmentation Method. In: Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition (2000)
- [9] Oliveira, L.S.: Automatic Recognition of Handwritten Numerical Strings. PhD thesis, Ecole de Technologie Superieure, Canada (2003)
- [10] Shubhangi, D.C., Hiremath, P.S.: Handwritten English Character And Digit Recognition Using Multiclass SVM Classifier And Using Structural Micro Features. International Journal of Recent Trends in Engineering 2, 193–195 (2009)
- [11] Verma, B.: Contour Code Feature Based Segmentation For Handwriting Recognition. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition (2003)
- [12] Xiao, X., Leedham, G.: Knowledge-based English cursive script segmentation. Pattern Recognition Letters 21, 945–954 (2000)
- [13] Yanikoglu, B., Sandon, P.A.: Segmentation of off-line cursive handwriting using linear programming. Pattern Recognition 31(12), 1825–1833 (1998)
- [14] Zhao, S., Chi, Z., Shi, P., Yan, H.: Two-stage segmentation of unconstrained handwritten Chinese characters. Pattern Recognition 36, 145–156 (2003)
- [15] Zheng, L., Hassin, A.H., Tang, X.: New algorithm for machine printed Arabic character segmentation. Pattern Recognition Letters 25, 1723–1729 (2004)

Part V
MEDICAL APPLICATIONS

Instantaneous Measurement of SNR in Electrocardiograms Based on Quasi-continuous Time-Scale Noise Modeling

Piotr Augustyniak

Abstract. Proper measurement of signal quality is a common problem in biomedical applications, including the electrocardiogram interpretation. The need for a reliable signal-to-noise estimate raises with the common use of telemedical recordings performed in the home care conditions and interpreted automatically. The currently used techniques perform noise measurements on the baseline and extrapolate the results to a whole heart beat. Main drawback of this method lies in irregular occurrence and short duration of the baseline. This paper presents a new ECG-dedicated noise estimation technique based on a time-frequency model of noise computed in a quasi-continuous way. The proposed algorithm uses the temporarily adapted local bandwidth variability of cardiac electrical representation to recognize cardiac components. The part of the time-frequency plane remaining above the local bandwidth of the signal, represents background activities of any origin (muscle, mains interference etc.). This noise estimate in each particular scale is available as non-uniformly sampled and has to be interpolated to the regions where components of cardiac representation are expected. In lower scales, the noise contribution is computed with use of square polynomial extrapolation. The algorithm was implemented and tested with use of the CSE Database records with the addition of the MIT-BIH Database noise patterns. The differences between the added and estimated noise show similar performance of baseline-based and noise model-based methods (0.69 dB and 0.64 dB respectively) as long as the noise level is stable. However when dynamic noise occurs, the baseline-based method is outperformed by the proposed algorithm (2.90 dB and 0.95 dB respectively) thanks to consideration of multiple measurement points and accurate noise tracking.

1 Introduction

The assessment of noise level in the electrocardiogram and - in a wider sense - the ECG record quality needs to be revised due to growing number of recordings

Piotr Augustyniak
AGH University of Science and Technology,
30 Mickiewicza Ave. 30-059 Kraków Poland
e-mail: august@agh.edu.pl

gathered in home care conditions. Unavoidable simultaneous activity of adjacent organs (e.g. muscles) and unstable recording conditions (e.g. electrodes contact) imply interferences with diagnostics or surveillance supportive signals. Both sources issue the noise of random occurrence overlapping the ECG signal in both time and frequency domains. Therefore, many noise measurement and removal techniques were developed in the past including signal averaging [12], adaptive noise canceling [1] or wavelet-based noise reduction [2], [10], [13], [14], [15]. These techniques, although relying on various conditions, are rarely applicable to home care recordings when the broadband noise contribution varies in energy. Moreover, most methods are based on the baseline as an interval of documented electrical inactivity of the heart and thus used as reference for noise measurement. Unfortunately, such approach has important limitations when applied to real ECG recordings: short duration of the baseline limiting the bandwidth, and rare, irregular occurrence of the baseline.

Our approach is based on the recognition of cardiac-originated components and background electrophysiological contributions performed in the time-frequency plane with use of the local bandwidth of the electrocardiogram (LBE). Time-frequency atoms found out of the ECG bandwidth are identified as noise measurement points and contribute to a quasi-continuous noise model in which interpolation or extrapolation is needed for only few gaps when cardiac representation fills all the time-frequency range. The relation of energy of cardiac components to the energy of the noise model is expressed directly in the time-frequency domain and yields a signal-to-noise ratio (SNR). Since the LBF is adapted to each consecutive heartbeat with use of individually detected wave borders (fig. 1a), the method shows a fast response to the local variability of background activity, compensates for variability of the heart rate and favorites the measured noise information over the estimates.

2 Recognition of Cardiac and Noise Components

The sampling frequency of 500 or 1000 Hz typical for ECG recordings, corresponds to relatively short QRS complex and is much too high for other cardiac components occupying the majority of recording time. The gap above the bandwidth expected for cardiac components and the Nyquist frequency is considered as temporal inactivity of the heart in specified frequency range (scale), similar to the baseline, but significantly longer (fig. 1b). These intervals are then used to measure the noise level at high frequency (three upper scales) directly on a scale-temporal plane. Because of different nature and expected bandwidth of each particular component of the heart evolution (the P, QRS and T waves), we found reasonable to correlate the local bandwidth estimate with positions of these waves, and not by the explicit value of time coordinate. Fortunately, the wave limits can be calculated automatically with acceptable reliability by the software. Consequently, the intervals of noise measurement are determined individually for each heart beat in three upper scales in intervals related to the content of the corresponding electrocardiogram. The local bandwidth of the ECG has been estimated in result of our previous research:

- studies of susceptibility of diagnostic parameter values to signal distortion caused by the local random canceling of time-frequency coefficients [4],
- analysis of expert's perception of the ECG trace revealing local signal conspicuity and thus its sectionwise relevance to the final diagnosis [5],
- individual assessment of the spectra for optimally-concatenated successive waves of selected types [8]

The heuristic function of local bandwidth expected at the time point n is expressed by a discrete function $f(n)$:

$$f : \forall n \in \{0, 1, \dots, N\} \rightarrow f(n) \in [0; 0.5] \quad (1)$$

representing the local relative cut-off frequency. This function, using $k_1 \dots k_5 \in \{0, 1, \dots, N\}$ as the representation of the standard positions of wave borders is projected to the local wave positions $h_1 \dots h_5 \in \{0, 1, \dots, M\}$ in the current heartbeat for each point $i = 1 \dots 5$ (fig. 1):

$$\forall n \in [k_i, k_{i+1}], \forall m \in [h_i, h_{i+1}] f'(m) = P^S(f(n)) \quad (2)$$

with projection scale S_i varying from section to section:

$$S_i = \frac{h_{i+1} - h_i}{k_{i+1} - k_i} \quad (3)$$

Time-frequency atoms of raw ECG representation are qualified as cardiac components only for scale j and time point m satisfying: $f'(m) > 2^{-j-1}$. Otherwise they are considered as extra-cardiac components (noise representation). Thanks to this approach, the model of ECG background activity contains as many measured data points as possible, without a risk of confusions of the noise with the cardiac representation. For each scale, the time-frequency noise model can be updated accurately without delay unless the scale contains the cardiac representation.

3 Continuous Estimate of Noise Contribution

In the sections containing atoms of cardiac activity, or where such activity is expected, the contribution of the noise has to be estimated from the previous and past measured values. For this purpose, consecutive values of noise measurement points in separate octaves N_j , $j \in \{1 \dots 3\}$, are considered as non-uniformly sampled time series $N_j(\{n, v(n)\})$ and projected to the regular space [3] using the continuous function:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (4)$$

where $x \in [x_i, x_{i+1}]$, $i \in \{0, 1, \dots, n-1\}$ best fitted to the time series N_j are known as cubic splines interpolation. The uniform representation of the noise, extended to the cardiac component area in three upper scales, is then obtained by sampling the $S_i(x)$ at the time points m (fig. 2a):

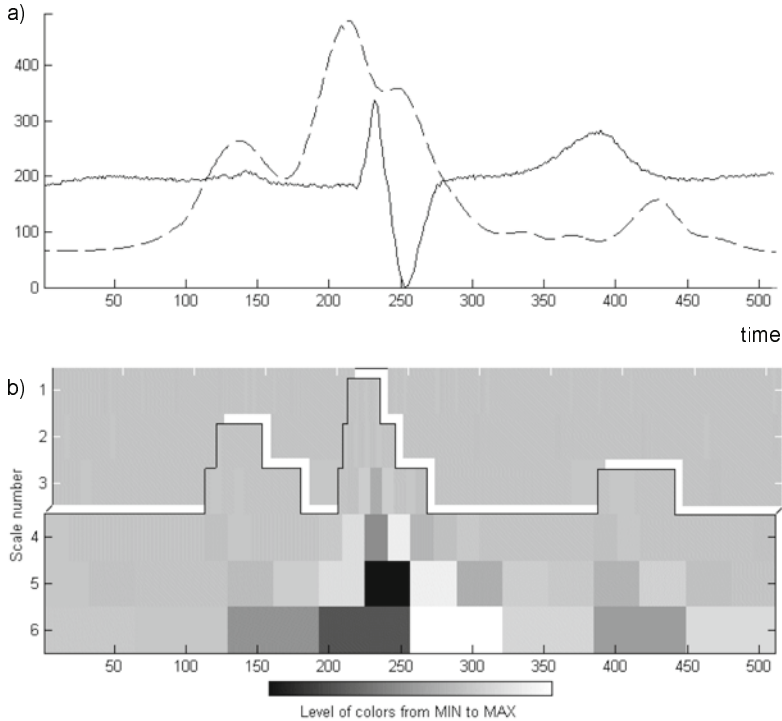


Fig. 1 a) The example heartbeat (solid) and the adapted bandwidth variability function (dashed). b) Corresponding time-frequency signal representation divided in the noise measurement region (above the local cut-off frequency) and the cardiac representation region (below)

$$N'_i(m) = \sum_m S_i(x) \cdot \delta(x - mT) \quad (5)$$

As the scale number increases, the contribution of cardiac representation grows and below 32 Hz ($j > 3$), reliable measurement of noise is never possible since the bandwidth is continuously occupied by the representation of cardiac activity. Therefore a noise extrapolation based on coefficients from first three scales is used to estimate the noise contribution in lower frequencies. This extrapolation uses second-order polynomials generated by all atoms of embedded trees originating from the considered coefficient:

$$N'(k, j) = a_{k,j} \cdot j^2 + b_{k,j} \cdot j + c_{k,j} \quad (6)$$

Therefore, the estimation of the noise level at a given time point k on the scale j is based on three average values $M_j(k, i)$ from all corresponding atoms $s(n, i)$ on each of the first three scales (fig. 2b):

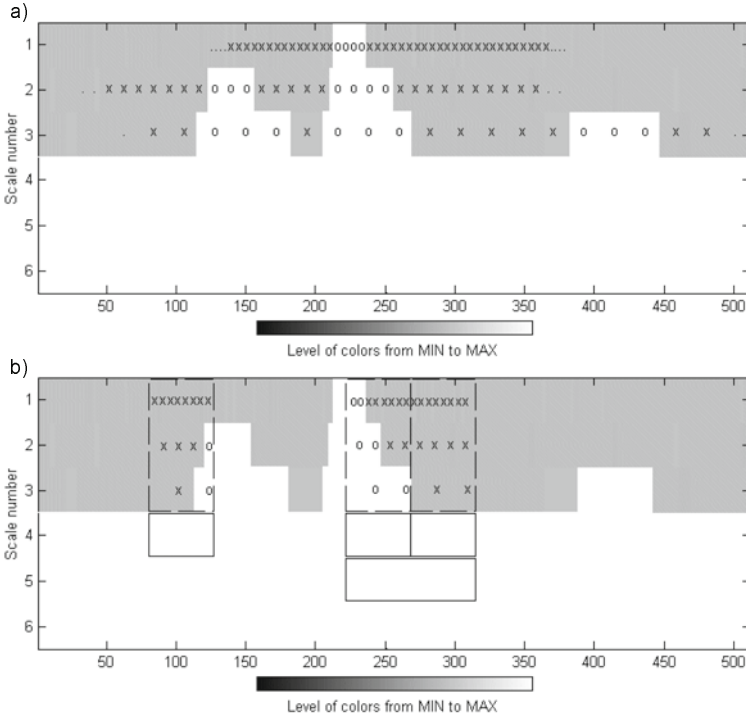


Fig. 2 (a) Distribution of noise measurement and interpolation samples in first three scales. (b) Extrapolation of noise values to low frequency bands with averaging of the noise print in the time domain. Missing values 'o' are estimated from adjacent measured values 'x'

$$M_j(k, i) = \frac{1}{2^{j-i}} \sum_{n=2^{j-i}k}^{2^{j-i}(k+1)-1} s(n, i) \quad (7)$$

The use of parabola-based extrapolation is justified by the envelope of the muscle representation spectrum and a consequence of a trade-off between the computational complexity and the temporal stability of the noise model.

4 Tests and Results

Testing the noise removal procedure is a difficult task because of the lack of noise-free real signals while the model (i.e. synthesized) electrocardiogram, although noise-free, differ in diagnostically meaningful details from the real ones. For this reason we perform two stationary noise tests using real and artificial records. For verification of dynamic features of the quasi-continuous noise model, we performed additional two tests using artificial records and the noise modulated by a sine and

Table 1 The average noise measurement accuracy [%] for synthesized and CSE-originated signals for three patterns of static noise

noise pattern level [dB]	synthesized				CSE Database			
	-3	-7	-10	-13	-3	-7	-10	-13
poor electrode contact	8	45	57	58	6	35	53	46
electromagnetic interference	66	78.5	87	81	66	78.5	87	81
muscle fibrillation	80	93	92.9	93.4	72	88.5	90	83

square waves. In all cases, three noise patterns artificially added to the original ECG record were representative for the most frequent noise sources:

- poor electrode contact (mainly low frequency and abrupt baseline changes),
- electromagnetic interference (mains sinus wave, 60 Hz),
- muscle fibrillation (a clipping-free section was selected)

All patterns were taken from the MIT-BIH Database (12 bit, 360 Hz) [11], resampled and mixed with the signal at the following four levels of total energy: 5%, 10%, 20% and 50% (corresponding to -13 dB, -10 dB, -7 dB and -3 dB signal-to-noise ratio). The artificial records were acquired with use of an ECG recorder (12 bit, 500 Hz) from the test generator (PS450, Fluke) and the real ECG records were taken from the CSE Multilead Database (12 bit, 500 Hz) [16].

4.1 Static Noise Test

In case of mixing the static noise with artificial ECG's the performance of both baseline-based and proposed noise model-based SNR measurements are very high, mainly due to separate spectra of the signal and noise (tab. 1, fig. 3). The artificial ECG signals are synthesized mathematically in the generator's low-noise hardware with no measurable random component. The time-frequency representation of the ECG perfectly matches the expected local bandwidth variability. Therefore any difference of estimated and added noise levels is interpreted as SNR-estimator inaccuracy. The results for real database records show lower performance, due to noise recorded in the reference signals, having the characteristics very close to the noise added artificially. This test, however, shows the behavior of SNR estimators in presence of irregular duration of ECG sections and waves.

4.2 Dynamic Noise Test

Dynamic tests were performed with artificial ECG's in order to reveal the dynamic properties of the proposed method. The changes of muscular activity were simulated by the modulation of noise energy with a sine wave of frequency sweeping from 1 to 10 Hz. This helps avoiding any correlation between the local noise level and the ECG content. The interference of real muscular biopotentials does not start or end abruptly, because the muscular fibre's triggers are not perfectly synchronized,

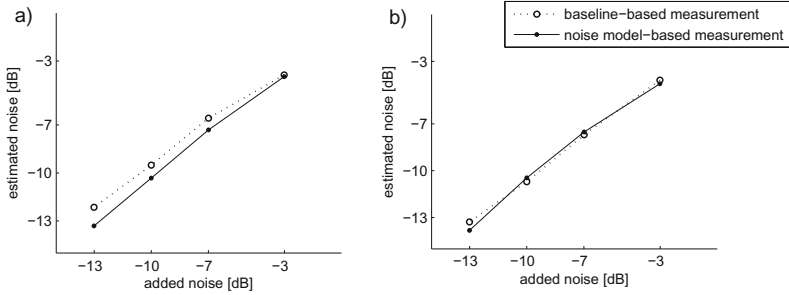


Fig. 3 SNR estimation performance for static muscular noise, plots show noise value estimated versus added (a) artificial signals, (b) CSE-originated signals.

Table 2 The average noise measurement accuracy [%] for synthesized signals for three patterns of sinus- and square modulated noise

noise pattern level [dB]	sinus modulation				square modulation			
	-3	-7	-10	-13	-3	-7	-10	-13
poor electrode contact	6	35	55	52	6	20	43	22
electromagnetic interference	66	78	86	78	54	54.5	56	62
muscle fibrillation	78	92	92.2	92.6	62	76.5	79	76

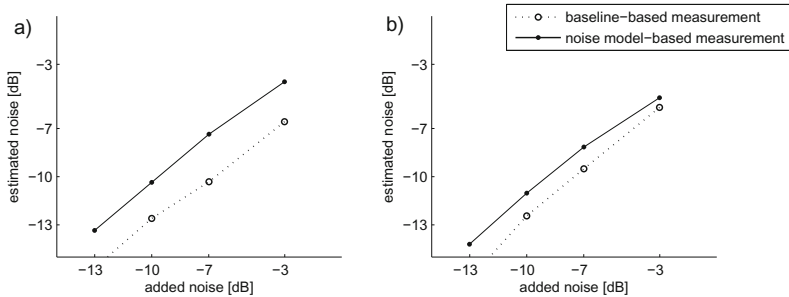


Fig. 4 SNR estimation performance for dynamic muscular noise, plots show noise value estimated versus added (a) sine wave modulation, (b) square wave modulation.

therefore the sine wave simulates well a natural appearance of noise. Finally we also performed this test with a square wave-modulated noise of frequency sweeping from 1 to 10 Hz, in order to determine the adaptability limits of the proposed time-frequency noise model (tab. 2, fig. 4).

In dynamic tests, in order to maintain the average noise level comparable with previous tests, the maximum amplitude of the mixed noise was twice as high as in case of continuous noise.

5 Discussion

The proposed algorithm for ECG noise recognition and calculation of signal-to-noise ratio is based on the definition of the local bandwidth of signal (i.e. cardiac components). This approach maximizes the number of noise measurement points in the time-frequency plane. Comparing to baseline-based SNR measurement using at most 10% of samples for noise level estimation, the proposed approach performs measurement on ca. 70% of the time-scale atoms, and interpolates only for remaining 30%. In consequence the noise tracking is more accurate and follows the rapid changes of the noise level, except if they occur in a QRS section (typically lasting for approximately 100 ms) where the full bandwidth of signal is used by the cardiac components. The measured noise data is then locally missing and the interpolation cannot correctly reconstruct the values of time-frequency coefficients. In practice this drawback has only marginal importance, because usually the amplitude of cardiac components on the QRS section is high and the signal masks the noise estimation errors.

Properties of the noise model-based SNR estimation are influenced by the definition of ECG local bandwidth, thus the result reported hereby corresponds to the average temporal distribution of diagnostic information in the discrete heartbeat representation. The use of local bandwidth function gives the user an opportunity to define the own profile of interest. The wideband approach defining the signal bandwidth as everywhere equal to the Nyquist frequency makes no assumption on the local signal properties, but also eliminates the points where the noise can be reliably measured. The opposite strategy defines the signal bandwidth to be much lower than the Nyquist frequency everywhere and allows the noise measurement at any time point like in oversampled signals. The appropriate bandwidth tracking enables the optimal measurement of the noise thanks to the use of general knowledge about the signal properties adapted to the current content of the data stream. These general remarks are valid for the wide class of physiological and technical signals characterized by irregular data distribution.

The noise patterns applied during tests differ in power spectra density and their characteristics influence the performance of the noise modeling and estimation algorithm. Since the poor electrode contact noise concentrates on lower scales, it cannot be correctly estimated using values measured from the three upper scales. Although tables 1 and 2 suggest a slight improvement in signal quality, the noise is canceled only in the middle frequency range (4th to 6th scales). The power line interference concentrates all the energy in very narrow frequency band influences the 2nd and 3rd scale (32...128 Hz) all the time. The presence of this time-invariant component lowers the sensitivity of noise adaptability in those scales and leads to the inaccurate frequency-domain estimation of the noise in lower scales. The characteristics of muscular activity best fit the initial assumptions: in the time domain signal changes are relatively slow and in the frequency domain the parabolic spectrum approximation is quite accurate. This is the reason for the best accuracy of the proposed algorithm in the tests with the muscle noise.

Since the time-frequency estimates of contribution from cardiac components and noise is expressed in energy units (i.e. square values of time-frequency coefficients), comparing their values calculated over all the signal duration yields directly the general and commonly used value of signal-to-noise ratio. However, considering a non-uniform distribution of medical information in the electrocardiogram, an interesting property of the proposed approach is the support for time-selective or event-specific assessment of the signal quality.

A new ECG-dedicated method for noise modeling and estimation was developed and tested. The noise model is quasi-continuous and optimizes the response to the physiological changes of background activity. The differences between the added and estimated noise show similar performance of baseline-based and noise model-based methods (0.69 dB and 0.64 dB respectively). However when dynamic noise occurs, the baseline-based method is outperformed by the proposed algorithm (2.90 dB and 0.95 dB respectively) thanks to maximum number of measurement points and accurate noise tracking.

Acknowledgment. This Scientific work is supported by the Polish State Committee for Scientific Research resources in years 2009-2012 as a research project No. N N518 426736.

References

- [1] Akay, M.: Biomedical signal processing. Academic Press, San Diego (1994)
- [2] Akay, M. (ed.): Wavelets in Biomedical Signal Processing. IEEE Press, New York (1998)
- [3] Aldroubi, A., Feichtinger, H.: Exact iterative reconstruction algorithm for multivariate irregularly sampled functions in spline-like spaces: the Lp theory. Proc. Amer. Math. Soc. 126(9), 2677–2686 (1998)
- [4] Augustyniak, P.: Controlling the Distortions Distribution in a Wavelet Packet-Based ECG Compression. In: International Conference on Image and Signal Processing, Agadir Morroco, pp. 267–277 (2001)
- [5] Augustyniak, P.: How a Human Perceives the Electrocardiogram. Computers in Cardiology 30, 601–604 (2003)
- [6] Augustyniak, P.: Time-frequency modelling and discrimination of noise in the electrocardiogram. Physiol. Meas. 24(3), 753–767 (2003)
- [7] Augustyniak, P.: Separating Cardiac and Muscular ECG Components Using Adaptive Modelling in Time-Frequency Domain. In: Proceedings of the WACBE World Congress on Bioengineering, paper 184 (2007)
- [8] Augustyniak, P.: Moving Window Signal Concatenation for Spectral Analysis of ECG Waves. Computing in Cardiology 37, 665–668 (2010)
- [9] Calderbank, A.R., Daubechies, I., Sweldens, W., Yeo, B.L.: Wavelet transforms that map integers to integers. Technical report, Department of Mathematics, Princeton University (1996)
- [10] Krishnan, S., Rangayyan, R.M.: Automatic de-noising of knee-joint vibration signals using adaptive time-frequency representations. Med. Biol. Eng. Comput. 38, 2–8 (2000)
- [11] Moody, G.B.: The MIT-BIH arrhythmia database CD-ROM, 3rd edn. Harvard-MIT Division of Health Sciences and Technology (1997)

- [12] Moss, A., Stern, S.: *Noninvasive Electrocardiology - clinical aspects of Holter monitoring*. Saunders Co., London (1996)
- [13] Nikolaev, N., Gotchev, A.: De-noising of ECG signals using wavelet shrinkage with time-frequency dependant threshold. In: *Proc. European Signal Processing Conf., EU-SIPCO 1998, Island of Rhodes, Greece*, pp. 2449–2453 (1998)
- [14] Nikolaev, N., Gotchev, A., Egiazarian, K., Nikolov, Z.: Suppression of electromyogram interference on the electrocardiogram by transform domain denoising. *Med. Biol. Eng. Comput.* 39, 649–655 (2001)
- [15] Paul, J., Reedy, M., Kumar, V.: A transform domain SVD filter for suppression of muscle noise artefacts in exercise ECG's. *IEEE Trans. Biomed. Eng.* 47, 645–662 (2000)
- [16] Willems, J.L.: *Common Standard for Quantitative Electrocardiography Multilead Atlas - Measurements results Data Set 3*. Commission of the European Communities - Medical and Public Health Research, Leuven (1988)

Data Queuing for Non-uniform Telemedical Reporting

Piotr Augustyniak

Abstract. Problems concerning the data queuing in non-uniformly reporting telemedical surveillance systems are investigated in this paper. Unlike the regular systems, where the data continuity is guaranteed by the common reporting interval and unified report content, the adaptive systems must implement a reservation procedure managing the content of every packet in order to proper data delivery, accordingly to sampling rates set individually for each of the diagnostic parameters. In the adaptive interpreting system, the reservation procedure has to consider changes in data flow caused by time-variable requirements for the update rate of the time series of particular diagnostic data. This can be achieved with consideration in the information structure of two auxiliary data parameters, the validity period and the priority. The proposed solution consists in two reporting modes: in *immediate mode* the diagnostic packets are transmitted immediately accordingly to the time requirements, while in *delayed mode* the transmission is deferred until packets are filled with valid data. Switching between these modes allows the telediagnostic system to respond in short time in case of emergency, and to limit the usage of data carrier for long-time regular reporting.

1 Introduction

The implementation of irregular processing and reporting in telemedical surveillance systems enhances the data communication effectiveness and helps avoiding of unnecessary energy dissipation. Although very promising [2], [3], this innovation implies problems with data uniformity that needs to be solved in order to guarantee the data continuity. The system considered as an example prototype is designed for cardiology-based surveillance of patients in motion and uses a star-shaped topology managed by the central server over the bi-directional digital data link [7], [9], [11],

Piotr Augustyniak

AGH University of Science and Technology, 30 Mickiewicza Ave.
30-059 Kraków Poland

e-mail: august@agh.edu.pl

[12], [15], [16]. The interpretation task is independent for each of the patients and shared between the remote recorder and the thread running on the central server. The interpretation of the recorded ECG signal and the adaptivity of diagnostic report content are subjects of multicriterial optimization performed by the server [5]. The interpretive software consists of four classes of procedures:

- recorder-only procedures, e.g. interfaces and resources management, signal acquisition and buffering, signal quality assessment,
- recorder defaults preloaded as dll's, e.g. QRS detection, morphology classification, wave delineation, arrhythmia detection,
- recorder optionals available as dll's, e.g. ST- and QT-segment analysis, HRV time domain analysis, pacemaker pulse detection,
- server-only procedures, e.g. HRV frequency domain analysis, VLP detection, HRT analysis and others.

Multiple versions of 2nd and 3rd class procedures are compiled for both: recorder and server platforms. The adaptation of the remote interpretation process is achieved in real time with the use of diagnostically-oriented libraries of subroutines. Subroutines within the libraries are commutable, thus designed for the same computational purpose, but with different prerequisites for the resources requirements and result quality. Their data gateways are standardized enabling the replacement required by the remote optimization, while the interpretation is running. Each subroutine is described by the attributes of quality, resources requirements and external dependency specifying its relations with the other components in the signal interpretation tree. Selected functions are uploaded from the repository managed by the supervising central server [1], [17] and dynamically linked with the software running in the patient's recorder. The adaptation is performed automatically, but occasionally, in critical cases, it is supervised by a human expert. Several variants of remote ECG interpretation and interactive adaptation of diagnostic report contents require exceptional flexibility of the data transmission format. This paper presents an effective solution for management of non-uniform flow of multimodal data, minimizing data redundancy but respecting individual update interval for each diagnostic parameter. It is based on dynamic definition of data structures and queuing of these structures in the output stream accordingly to the minimum value of update interval (validity time).

2 Report Data Structures

2.1 Flexible Reporting Format

As designed for the adaptive diagnostic system, the reporting format has to provide as much flexibility as possible. We implemented three-layer data format consisting of:

- mandatory header describing packets content and recorder's status,
- mandatory data description fields with pointers,
- optional data fields.

The header contains unique identifier of the recorder and the static structure of 18 bytes representing its status [2]. Next fields contain the session identifier, sequential packet number and the time interval to the subsequent report. Two later data are crucial for the correct report queuing and for the assessment of data integrity and continuity. Remaining header fields are used for description of the first data field type and contains the pointer to the first byte of its structure. First such field is mandatory even if the report is empty (i.e. contains only the end-of-file mark), but more often the descriptives of data fields with respective pointers are repeated within the header as many times as necessary. This allows each report to be individually composed of any combination of predefined data structures. For the reason of multitude of data meanings (including raw signals, meta data, diagnostic parameters) and forms (simple variables, vectors, matrices and structures) used for the reporting, the description of data organization precedes the report contents. The format design supports both predefined and ad-hoc defined data structures, however in course of implementation the use of the ad-hoc mode was found not efficient and discontinued. In the data description layer, each data type specification is followed by at least one data allocation triplet consisting of the index, length and pointer to the particular record serialized in the data layer. In case of several records complying with the common data type definition, all of them are integrated to a vector. Medical data serialized in their description order are the sole content of the data layer. First field is pointed to by the pointer of the first segment of first defined data type. It is mandatory, in case of empty report it contains the end-of-file mark. Minimum length of the diagnostic report is 20 bytes, while the maximum length is unlimited and depends only on the transmitted datastream. For practical reasons, reports below 256 bytes are sent only in immediate reporting mode and reports exceeding 1500 bytes are split due to the limitation from the TCP/IP specification [8], [10].

2.2 *Dynamic Definition of Data Structures*

The definition of data organization is a part of the adaptation process of patient's recorder software. Unlike the software modification, which may be executed in course of the patient monitoring, we assume to exchange the data structure description once at the beginning of each monitoring session. Up to 256 different data structures may be defined for each session, what was expected as sufficient for the necessary flexibility of the report content. Moreover, definitions of complex structures may embed the simpler, previously defined types. This allows for definitions of nested data structures (fig. 1) particularly efficient if the update interval of one data aggregate is a multiple of the other. The patient status determines the further record interpretation and the resulting parameters set. Each diagnostic parameter is attributed with the priority and the validity time, the latter being the inverse of the minimum update rate required to maintain the continuity of each time series.

Although reporting each diagnostic parameter with its proper update rate optimally reduces the medical data stream, embedding individual diagnostic values into the communication protocol control information raises the contribution of the

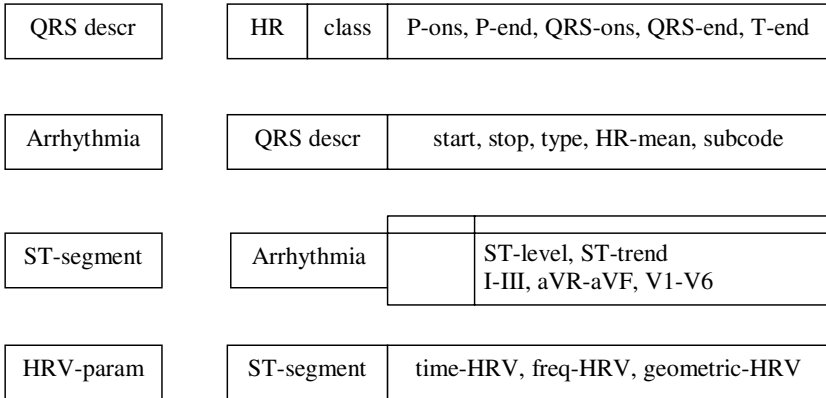


Fig. 1 Design of sample information structures for the adaptive reporting in the prototype of seamless cardiac supervising system

TCP/IP overhead. In order to maintain the communication effectiveness, it was reasonable to aggregate the independent data having similar validity time into a common structure. Its values are then updated accordingly to the update rate of the most varying parameter causing only a slight oversampling of the others. Each time, the interpretation software is modified, the report reservation procedure matches the attributes of diagnostic parameters with the information structures available for the session:

- Parameters are clustered by their value of required update interval and each class is assigned by the structure best reflecting its content.

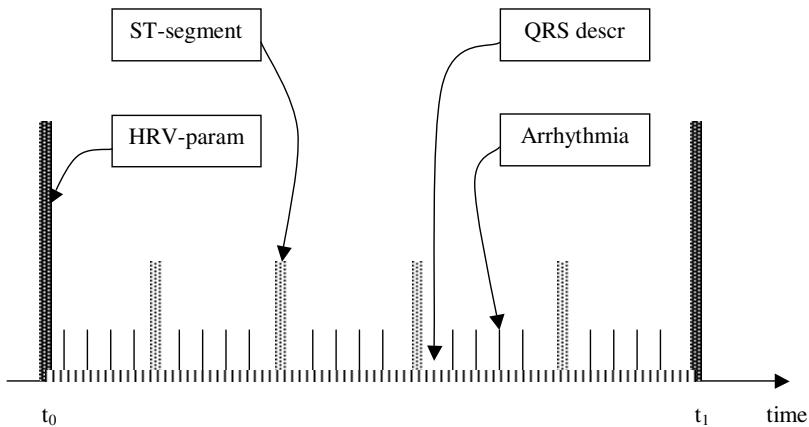


Fig. 2 Sample sequence of medical data structures used for reporting of the heart rate, arrhythmia events and level of ST section

- A sequence of structures is defined and scheduled accordingly to the diagnostic parameter of shorter validity time.

The cluster size, measured as the difference of required update frequencies between its border members, is adjusted with regard to the volume of TCP/IP control data. The proposed data aggregation and scheduling rule guarantees the reporting continuity of each parameter and maintains the global data on its lowest possible volume. In practical experiments aimed at testing the adaptivity of the reports, the number of necessary different data types rarely exceeded 16.

3 Review of Reporting Modes

3.1 *Delayed Reporting*

Delayed reporting gathers the data accordingly to the required update rate, but allows for collecting the consecutive values in a single report in case the data volume is too low to fill the entire packet. This saves the protocol control data volume and the transmission energy at the price of delay in information delivery. This mode is defined for the routine monitoring of stable subjects. Due to data buffering, the transmission module may be temporarily switched to the low-power state in order to extend the recorder's autonomy time. The diagnostic patient's representation received by the server is continuous, however due to the delay, the last available data may not be recent. The *delayed reporting* doesn't require the continuous availability of wireless data carrier. If the data carrier is available, the *delayed reporting* may be automatically switched to the *immediate reporting* mode in following cases:

- Occurrence of pathological events,
- Activation of the patient's button,
- Adaptation of the remote recorder interpreting software,
- Recovering from the carrier absence, if originally the *immediate* mode was set.

In the prototype implementation, the *delayed reporting* mode buffers the subsequent diagnostic reports up to the size of 32kB or up to the delay of 360s. If one of these limits is reached, the transmission module is activated and the packet containing several consecutive diagnostic reports is sent.

3.2 *Immediate Reporting*

In the *immediate reporting* mode the report is sent as soon as all included diagnostic parameters are calculated. This mode maintains both the continuity and minimum delay of the diagnostic patient's representation available at the server. This mode is particularly useful for the monitoring of high-risk patients, for synchronization of the patient-doctor interaction or for the immediate assessment of adaptation of the recorder's interpretation software. This mode requires the continuous operation of the transmission module and availability of wireless data carrier. The *immediate*

reporting may be automatically switched to the *delayed reporting* mode in following cases:

- Wireless data carrier absence,
- Recession of any condition imposing immediate reporting mode, if originally the *delayed mode* was set.

4 Management of the Reports Content and Frequency

The adaptive nature of ECG interpreting software implemented in the remote (patient) recorder implies non-uniform content of the diagnostic report. Moreover, depending on the patient's status, the update interval attribute is individually set by the management software. For those reasons, the management of the reports content and frequency is directly related to the software adaptation.

The procedure managing of the data queuing in adaptive diagnostic reports is implemented in the remote recorder and - in spite of the importance of the performed optimization - should not require much of additional complexity. When the adaptation of the interpretation software is completed, the combination, the priority and the validity intervals of diagnostic parameters has to be analyzed. As the result, most appropriate data structures and their repetition intervals are determined as components used for building the report sequence. The sequence starts with the structure composed of all considered diagnostic parameters (ref. fig. 2 point t_0) and then the subsequent structures and intervals are scheduled with regard to the required update interval. In case when several structures are applicable for a given parameters combination, the structure of minimum size is selected. When the use of the structure composed of all considered parameters is necessary again (point t_1), the queuing procedure assumes the report sequence is completed and contains the basic repetition period. The sequence is used for the reporting until the next adaptation of the recorder's interpretive software.

5 Results

Before the implementation in the target operating system [19], the prototype data queuing procedure was implemented in Matlab. This implementation was aimed at testing the automatic building of report sequences for various combination of diagnostic parameters and their update intervals.

5.1 Testing Conditions

Four basic data structures were created as default for the transmission (see fig. 1):

- heart beat descriptor *QRS Descr* (volume: 16 bytes, update interval: 300 ms)
- arrhythmia descriptor *Arrhythmia* (volume: 5 bytes, update interval: 3 s)
- ST-segment descriptor *ST-segment* (volume: 24 bytes, update interval: 30 s)

- HRV parameters descriptor *HRV-param* (volume: 10 bytes, update interval: 360 s)

Since the test was focused on the dynamic behavior of data queuing, and cardiac changes in real patients are unexpected, we used 2751 signals artificially combined from 58 records of Common Standards for quantitative Electrocardiography (CSE) Multilead Database [18]. In each test signal the appearance of a single pathology was simulated by the concatenation of several repetitions of a record annotated as normal followed by several repetitions of a pathology-specific record. In the vicinity of the concatenation point, the count of samples was adjusted accordingly to the precedent RR interval and the baseline level of the subsequent record was corrected for each lead independently. Such test records simulate the abrupt cardiac changes and facilitate the technical analysis of the adaptive system behavior.

5.2 Test Results

Main results presented in table 1. show data volume only for the reports composed of few basic diagnostic parameters and their most probable update intervals. These results differ accordingly to the patient status, since various pathologies require specific diagnostic data set and assume individual variability of these data implying variations of update intervals. In the *delayed reporting* mode, the size limit of 32kB requires 580s of recording. This value corresponding to maximum reporting delay of 9 min 40 seconds is high even for the stable subject. Therefore the data buffering first achieves the interval limit of 360s (6 minutes), in this time, the buffer accumulates nearly 20kB of data. The wireless telecommunication module in a target hardware platform remains operating (supply current 50mA @ 5V) for 6 seconds before the data transfer starts in order to re-establish the connection. The transfer session lasts for ca. 4 to 10 seconds depending on the speed of data transfer (16 to 40kbps). Including additional 2 seconds for the session termination, the module is operating for 18s out of 360s reporting interval. Since the energy consumption in standby mode is almost negligible (0.5 mA), the average energy in *delayed reporting* is reduced below 3mA (i.e. to 6% of the original value). In the *immediate*

Table 1 Comparison of the data stream volume [bps] using the constant reporting rate and the adaptive reporting with data queuing method proposed in this paper (12-lead ECG)

component	volume [b]	interval [s]	theoretical data stream	adaptive rate data stream	constant rate data stream
heart rate	1	0.3	55,277	65,67	181,5
morphology	1	0.3			
wave borders	14	0.3			
arrhythmia events	5	3			
ST-segment elevation	24	30			
HRV parameters	10	360			

reporting mode, the minimum packet size of 256 bytes is collected within 4,63 seconds, which is slightly worse than the delay in the interpretation process (typically 2s), but still acceptable as real time monitoring.

6 Discussion

An effective solution was proposed for the patient's recorder data queuing in a non-uniformly reporting cardiac surveillance system. For each adaptively selected combination of diagnostic parameters requiring specification of individual time-varying update rate, the automatic procedure schedules the data structure sequence of minimum size, maintaining the reporting continuity. This procedure is designed for the cooperation with software adaptation process, but may also be used as a software interpretation trigger in "on demand" interpretive systems [4], [6].

As it has been demonstrated in tab. 1, the regular ECG reporting implies a significant oversampling of certain parameters, and consequently - the waste of energy and telecommunication resources. The intelligent data queuing requires the use of various data structures and individual management of packet delivery. Each data being independently sent, involves the use of TCP/IP control information, therefore high fragmentation of diagnostic reports, although enables the parameters bandwidth variability to be precisely followed, increases the total volume of the transferred data. The most reasonable compromise consists in building the compound information structures of diagnostic data requiring similar update interval. Although some parameters are reported more frequently than it is necessary, such solution helps to maintain minimum data volume thanks to using larger blocks of data and lower contribution from the protocol control information.

The optimization of the contents of diagnostic report is an example of pursuit for a local minimum of a cost function (in this particular case represented by the volume of redundant data) in a multidimensional discrete space with constrains. These constrains are:

- requirement (absolute) of minimum updating frequency set for each diagnostic parameter individually,
- condition (relative) of usable volume of diagnostic report,
- condition (relative) of minimum number of data structure definitions.

Although the proposed non-uniform data management method assumes unprecedented flexibility, it does not interfere with any existing data transfer standard (e.g. SCP-ECG [13]). Thanks to a general definition background it extends the support of existing standards to non-uniform time series of diagnostic data, being a natural way of reporting from an adaptive interpreting software. This solution is thus a step towards the remotely personalized medicine.

Acknowledgment. This Scientific work is supported by the Polish State Committee for Scientific Research resources in years 2009-2012 as a research project No. N N518 426736.

References

- [1] Atoui, H., Telisson, D., Fayn, J., Rubel, P.: Ambient Intelligence and Pervasive Architecture Designed within the EPI-MEDICS Personal ECG Monitor. *International Journal of Healthcare Information Systems and Informatics* 3(4), 68–80 (2008)
- [2] Augustyniak, P.: Content-Adaptive Signal and Data in Pervasive Cardiac Monitoring. *Computers in Cardiology* 32, 825–828 (2005)
- [3] Augustyniak, P.: Task-dependent adaptability of remote recorder in a cardiac surveillance network. In: *World Congress on Bioengineering*, pp. 1–4 (2007)
- [4] Augustyniak, P.: Request-driven ECG interpretation based on individual data validity periods. In: *Proc. 30th Conf. IEEE Engineering in Medicine and Biology Society*, pp. 3777–3780 (2007)
- [5] Augustyniak, P.: Validation of automatic ECG processing management in adaptive distributed surveillance system. In: Kurzyński, M., et al. (eds.) *Computer Recognition Systems 2*, pp. 574–580. Springer, Heidelberg (2007)
- [6] Augustyniak, P.: Strategies of Software Adaptation in Home Care Systems. In: Kurzyński, M., Woźniak, M. (eds.) *Computer Recognition Systems 3. Advances in Intelligent and Soft Computing*, vol. 57, pp. 393–400. Springer, Heidelberg (2009)
- [7] Balasz, G., Kozmann, G., Vassanyi, I.: Intelligent Cardiac Telemonitoring System. *Computers in Cardiology* 31, 745–748 (2004)
- [8] Banitsas, K.A., Georgiadis, P., Tachakra, S., Cavouras, D.: Using handheld devices for real-time wireless Teleconsultation. In: *Proc. 26th Conf. IEEE Engineering in Medicine and Biology Society*, pp. 3105–3108 (2004)
- [9] Boussejot, R., et al.: Telemetric ECG Diagnosis Follow-Up. *Computers in Cardiology* 30, 121–124 (2003)
- [10] Chen, X., Ho, C.T., Lim, E.T., Kyaw, T.Z.: Cellular Phone Based Online ECG Processing for Ambulatory and Continuous Detection. *Computers in Cardiology* 34, 653–656 (2007)
- [11] Chiarugi, F., et al.: Continuous ECG Monitoring in the Management of Pre-Hospital Health Emergencies. *Computers in Cardiology* 30, 205–208 (2003)
- [12] Fayn, J., et al.: Towards New Integrated Information and Communication Infrastructures in E-Health. Examples from Cardiology. *Computers in Cardiology* 30, 113–116 (2003)
- [13] Fischer, R., Zywiets, C.: How to implement SCP (2001), retrieved from <http://www.openecg.net>
- [14] Paoletti, M., Marchesi, C.: Low computational cost algorithms for portable ECG monitoring units. In: *IFMBE Proc. Medicon*, paper 231 (2004)
- [15] Pinna, G.D., Maestri, R., Gobbi, E., La Rovere, M.T., Scanferlato, J.L.: Home Telemonitoring of Chronic Heart Failure Patients: Novel System Architecture of the Home or Hospital in Heart Failure Study. *Computers in Cardiology* 30, 105–108 (2003)
- [16] Puzzuoli, S.: Remote Transmission and Analysis of Signals from Wearable Devices in Sleep Disorders Evaluation. *Computers in Cardiology* 32, 53–56 (2005)
- [17] Telisson, D., Fayn, J., Rubel, P.: Design of a Tele-Expertise Architecture Adapted to Pervasive Multi-Actor Environments. Application to eCardiology. *Computers in Cardiology* 31, 749–752 (2004)
- [18] Willems, J.L.: *Common Standards for Quantitative Electrocardiography 10-th CSE Progress Report*. ACCO publ., Leuven (1990)
- [19] Ziecik, P.: Mobile Development Platform for Wide Area ECG Monitoring System. In: *Proc. of Biosignal 2008 Conference*, paper 72 (2008)

Analysis of Dispersion of Isopotential Areas Values in EEG Maps Sequences*

Hanna Goszczynska, Marek Doros, Leszek Kowalczyk, Paweł Hoser,
and Krystyna Kolebska

Abstract. The aim of the study was to develop the method for objective evaluation of differences in alternation variability of isopotential areas in the EEG maps sequences for activity episode. The method was based on the analysis of EEG map images histograms and concerns the evaluation of the isopotential areas changes. The principle of the present method is analysis of the ratio of extreme isopotential areas and of the dispersion of values of extreme isopotential areas. The method was tested on sequences of amplitude maps for 17 subjects with seizure episodes with different characteristics divided into two groups. Visual evaluation of obtained results confirms that ratio of extreme isopotential areas and coefficients of the symmetry and elongation of extreme isopotential areas values dispersions may be useful for describing the variability of EEG maps. Influence of the potential range scale of EEG maps on the quantitative estimation of the isopotential areas variability was also analyzed.

1 Introduction

Brain Electrical Activity Mapping – BEAM (*Duffy FH, Burchfield JL, Lombroso CT: Brain electrical activity mapping (BEAM): a method for extending the clinical utility of EEG and evoked potential data. Ann Neurol 1979,5:309-321*) is a well established electroencephalographic (EEG) examinations method as utilizing imaging of the distribution of different values (parameters) characterizing the bioelectrical brain activity measured on the skull surface [1, 9, 12]. In the development of the functional neuroimaging the different forms of the diagnostic images arose, e.g. momentary maps, instantaneous maps, significant probability mapping [2], differential maps [16], landscapes [8]. However, they are not commonly applied in practice.

Hanna Goszczynska · Marek Doros · Leszek Kowalczyk · Paweł Hoser · Krystyna Kolebska
M. Nałęcz Institute of Biocybernetics and Biomedical Engineering PAS, 4,
Trojdena str., 02-109 Warsaw
e-mail: hgoszczynska@ibib.waw.pl

* This work was partially supported by the Ministry of Science and Higher Education grant no 3291/B/T02/2008/35.

Mapping of brain electrical amplitude is a routine method used in electroencephalography. Maps sequences are very useful for presentation of the variability of constellations depicted on maps. This may be of significance in the evaluation of different episode like eg. seizure activity and its dynamics. Evaluation of the consecutive maps changes requires, however, the quantitative methods [10, 11, 13, 14, 15].

The aim of the present as well as previous study [3, 4] was to develop the methods for objective evaluation of differences in alternation variability of isopotential areas in the EEG maps sequences with seizure activity episode. Alternation of isopotential areas is manifested by changes of isopotential areas values as well as changes of topologicalisation of these areas. In clinical practice this variability is evaluated using visual inspection that is subjective and may be difficult in case of subtle differences. Quantitative analysis of isopotential areas variability could be useful for evaluation of brain electrical activity dynamic, e.g. propagation of seizure activity or reactivity to stimulus. Our approach has been based on the analysis of EEG map images histograms and concerns the evaluation of the change of isopotential areas values. In this paper application of standard measures for evaluation of the dispersion of the values of isopotential areas is presented. Influence of the potential range scale of EEG maps on the quantitative estimation of the isopotential areas variability was also analyzed.

2 Material and Method

The material comprised 17 subjects divided into two groups with episodes of different characteristics classified in previous works [5, 6, 7]. There is common numbering for both groups of subjects: 1 - 7 for group II and 8 - 17 for group I. For each subject the sequence of amplitude maps generated with frequency equals to 100 Hz for ten seconds of EEG records consisting of pre-, during and post- episode activity has been collected. Each map was generated in 17 levels color scale corresponding to 17 ranges of potential values $S1 - S17$ equal to $\pm 20 \mu V$ [4]. $S1$ denotes minimum potential range and $S17$ denotes maximum potential range. Respective isopotential areas are denoted as A_{S1} and A_{S17} .

Fig. 1a and Fig. 1d present the consecutive 50 frames of EEG maps at the beginning of the episode for subjects no. 3 from the group II and no. 9 from the group I, respectively.

The Fig. 1a presents the sequence of maps with lower variability of the areas values which were distinctive for group II, while the maps in group I presented higher variability (Fig. 1d). Using the normalized histograms (Fig. 1b and Fig. 1e) for all frames in the given sequence the values of areas for amplitude ranges $S1$ and $S17$ were calculated. For assumed scale $\pm 20 \mu V$ respective isopotential areas are denoted as A_{-20} and A_{20} . The principle of the method for the variability of the areas A_{-20} and A_{20} estimation in EEG maps sequence consist of:

- analysis of the plot of ratio of time series $A_{-20}(t)$ and $A_{20}(t)$ for the areas values variability evaluation presented in [3, 4],

- analysis of the scatter plot of points $p(A_{-20}(t), A_{20}(t))$ (Fig. 1c and Fig. 1f). The quantitative analysis of the points $p(A_{-20}, A_{20})$ dispersions was performed using the standard measures of distribution of time series $A_{-20}(t)$ and $A_{20}(t)$ and comparing the results of distribution measures for both time series. Also the analysis of the points $p(A_{-20}, A_{20})$ dispersions was performed by calculating the eigenvalues of dispersion matrix for both time series.

Also the criteria for the analysis of the influence of maps scale on the dispersion of points $p(A_{-20}, A_{20})$ and on the quantitative coefficients differentiating the configuration of points $p(A_{-20}, A_{20})$ are presented.

The analysis of distribution of time series $A_{-20}(t)$ and $A_{20}(t)$ has been applied for estimation of the points $p(A_{-20}, A_{20})$ dispersion. Two time series for each sequence of 1000 maps comprising of periods before, during and after the seizure episode activity have been analyzed. Different time of duration of episode in each maps sequence makes the dispersion measures of isopotential areas time series incomparable for different subjects. Only the ratios of the distribution measures for both time series were analyzed for estimation of their similarity and therefore the symmetry of points $p(A_{-20}, A_{20})$ dispersion.

Two coefficients of distribution for time series $A_{-20}(t)$ and $A_{20}(t)$ have been calculated:

- normalized moment of 3rd order a_3 describing the skewness of distribution,
- normalized moment of 4th order a_4 describing the kurtosis of distribution.

As a measure of the points $p(A_{-20}, A_{20})$ scatter plot elongation the eigenvalues of dispersion matrix of time series $A_{-20}(t)$ and $A_{20}(t)$ have been calculated and the

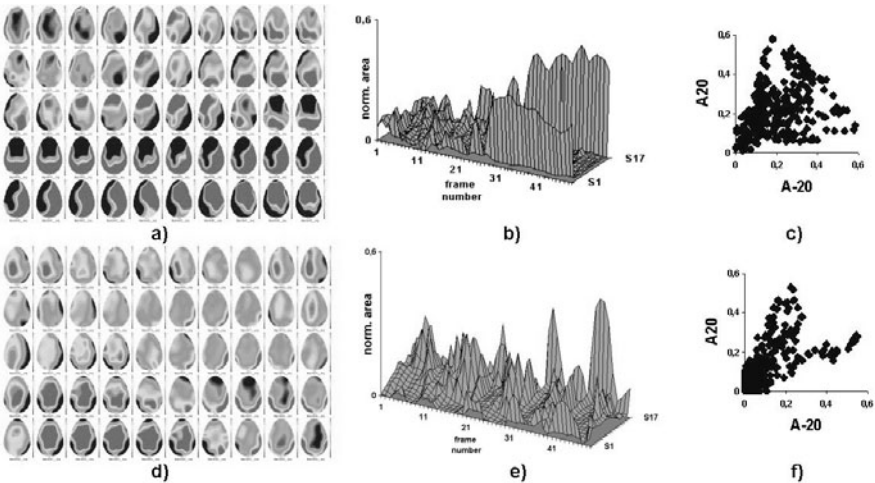


Fig. 1 Representative sequences of 50 frames of EEG maps containing beginning of episode (a,d), sets of its normalized histograms in 3D form (b,e) and scatter plots of the points $p(A_{-20}, A_{20})$ (c,f) for subject no.3 and no. 9, respectively

differences of the values of maximal and minimal eigenvalues of dispersion matrix has been assumed as a measure of scatter plot elongation. Dispersion matrix W_A of time series is defined according to the formula [17]:

$$W_A = L^{-1} \sum_{i=1}^L (p_i - p^h) \otimes (p_i - p^h) \quad (1)$$

where W_A – dispersion matrix, p_i – successive point $p(A_{-20}, A_{20})$, p^h – central point of a set of the analyzed points, L – number of analyzed points. Maximal and minimal eigenvalues of matrix W_A describes the dispersion of points as the axis of ellipse fitted to the set of points.

The amplitude range of individual EEG records during seizure episode for both groups of subjects was from $\pm 30\mu\text{V}$ to $\pm 200\mu\text{V}$. All maps were generated with constant scale $\pm 5\mu\text{V}$ equals to $\pm 20\mu\text{V}$. After the visual analysis of EEG records for both groups of subjects the scales for EEG maps generation for selected records have been changed and measurements of isopotential areas and calculations of the distribution coefficients have been repeated.

The following criteria of EEG records selection to respective maximal amplitude have been given:

- records for which the scale $\pm 20\mu\text{V}$ seem to be too narrow,
- segments, for which the scale $\pm 20\mu\text{V}$ seem to be too wide.

For selected records the sequences of EEG maps in various scales have been generated and quantitative coefficients of points $p(A_{-S1}, A_{S17})$ dispersion have been calculated.

3 Results

Fig. 1c and Fig. 1f present scatter plots for points $p(A_{-20}, A_{20})$ for subjects from both group of subjects. A visual evaluation of scatter plots for isopotential areas A_{-20} and A_{20} for both groups of subjects allowed to distinguish the following characteristic configurations: "dispersive" (Fig. 1c) present in II group and "bi-tailed" (Fig. 1f) present in I group (about 50% cases). The remaining 50% of cases contained the features of the two mentioned configurations.

Quantitative estimation of points $p(A_{-20}, A_{20})$ dispersion were performed by analysis of the similarity of the distribution of time series $A_{-20}(t)$ and $A_{20}(t)$ using skewness and kurtosis coefficients and analyzing the elongation of points $p(A_{-20}, A_{20})$ scatter plot using the differences of eigenvalues of the dispersion matrix for the time series $A_{-20}(t)$ and $A_{20}(t)$.

The ratios of skewness coefficients and the ratios of kurtosis coefficients for time series $A_{-20}(t)$ and $A_{20}(t)$ for both groups of subjects are presented in Fig. 2. The values of skewness and kurtosis for both time series and eigenvalues for dispersion matrix are shown in upper part of Fig. 2.

The ratios of skewness (Fig. 2a) and kurtosis (Fig. 2b) for the group II are about 1. It indicates the similar (in the sense of asymmetry and dispersion around the

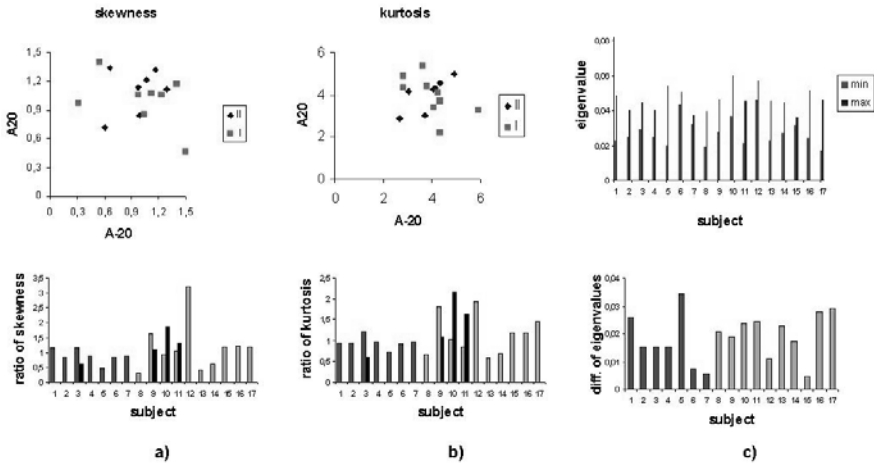


Fig. 2 Values of ratios of skewnees (a) and kurtosis (b) coefficients for time series $A_{-20}(t)$ and $A_{20}(t)$ and values of differences of eigenvalues for the dispersion matrix (c) for the time series $A_{-20}(t)$ and $A_{20}(t)$ for both groups of subjects

average) distributions of series $A_{-20}(t)$ and $A_{20}(t)$ for group II, while for the group I the values of the respective ratios indicate the differences in distributions of series $A_{-20}(t)$ and $A_{20}(t)$.

Fig. 2c displays the diagrams of values of dispersion measure defined as the difference of eigenvalues of the dispersion matrix of the time series $A_{-20}(t)$ and $A_{20}(t)$. Values of the eigenvalues differences of the dispersion matrix are smaller in group II (except subjects no. 1 and no. 5) than in group I (except subjects no. 12 and no. 15). It indicated for more elongated configuration of points $p(A_{-20}, A_{20})$ in group I than in group II.

Visual evaluation of presented results confirms that examined coefficients may be useful for describing the dispersion of the points $p(A_{-20}, A_{20})$: ratios of skewness coefficients and kurtosis coefficients for the time series $A_{-20}(t)$ and $A_{20}(t)$ and the difference of eigenvalues of the dispersion matrix for the time series $A_{-20}(t)$ and $A_{20}(t)$.

The influence of the scale for which the EEG maps are generated has been tested on the selected data. Two subjects for whom the scale $\pm 20\mu V$ seemed to be underestimated: patient no. 3 with maximal amplitude equal to about $200\mu V$ and subject no. 9 with maximal amplitude equal to about $80\mu V$ have been selected. For these cases new scale equal to $\pm 40\mu V$ have been assumed. Fig. 3a presents the consecutive 50 frames of EEG maps at the beginning of episode for subject no. 3 and the 3D view of sets of their normalized histograms (Fig. 3b) (comp. with Fig. 1b) and the scatter plot of points $p(A_{-40}, A_{40})$ (Fig. 3c). Two subjects for which the scale $\pm 20\mu V$ seemed to be overestimated: subjects no. 10 and no. 11 with maximal amplitude equal to around $30\mu V$ have been selected. For these cases new scale equal to $\pm 15\mu V$ has been assumed.

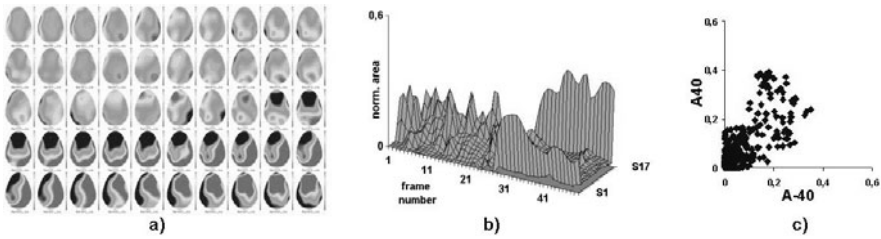


Fig. 3 Sequences of 50 frames of EEG maps in scale $\pm 40\mu\text{V}$ (a), sets of their normalized histograms (b) and scatter plot of the points $p(A_{-40}, A_{40})$ for subject no. 3 (comp. Fig. 2a-c)

A visual evaluation of the configuration of the points $p(A_{S1}, A_{S17})$ dispersion for selected four cases indicated the reason of change of the range of potentials from $\pm 20\mu\text{V}$ to $\pm 15\mu\text{V}$ (subject no. 10 and no. 11). On the contrary, the change of range of potentials from $\pm 20\mu\text{V}$ to $\pm 40\mu\text{V}$ (subjects no. 3 and no. 9) causes decrease the number of points in scatter plots and therefore loss of significant information (Fig. 3c).

The values of ratio of skewness coefficients and kurtosis coefficients for time series $A_{S1}(t), A_{S17}(t)$ for four selected subjects were presented in Fig. 2a and Fig. 2b marked by black lines. Change of the scale to $\pm 15\mu\text{V}$ for the subjects no. 10 and no. 11 caused the shift of values of ratio of skewness and kurtosis to the direction prevailing in group I. On the contrary, change of the scale to $\pm 40\mu\text{V}$ for the subjects no. 3 and no. 9 caused the shift of values of coefficients to the direction opposite to prevailing in given groups.

4 Discussion

Presented method allows to estimate one of the alternation of isopotential areas features, i.e. isopotential area values variability. The results of variability estimation obtained in present as well as previous [3, 4] study were affected by the following factors:

- analysis of the change of isopotential areas comprises the "pathological" as well as "physiological" fields,
- a topolocalisation changes were not taken into consideration,
- the scale of generated maps was constant.

The visual evaluation of the scatter plots presented in Fig. 1 indicates the rather regular dispersion of points $p(A_{-20}, A_{20})$ for group II. For group I there were two clusters for points with coordinates higher than assumed threshold. Configurations of points $p(A_{-20}, A_{20})$ dispersion regard estimation of dynamic of the ratio $A_{-20}(t)/A_{20}(t)$ [3]: lower dynamic of the ratio $A_{-20}(t)/A_{20}(t)$ (Fig. 4a) regards the scatter configuration of points p for group II and higher dynamic of the ratio $A_{-20}(t)/A_{20}(t)$ (Fig. 4b) regards bi-tail configuration of points p for group I.

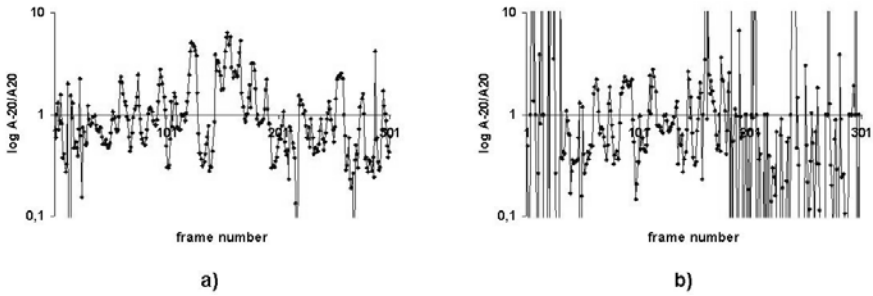


Fig. 4 Values of ratios of time series $A_{-20}(t)$ and $A_{20}(t)$ in logarithmic scale for subjects no. 3 (a) and no. 9 (b) for 300 frames (first 50 frames presented in Fig. 1)

Quantitative estimation of points $p(A_{-20}, A_{20})$ dispersion were performed by analysis of the similarity of the distribution of time series $A_{-20}(t)$ and $A_{20}(t)$ using skewness and kurtosis coefficients and analyzing the level of elongation of points $p(A_{-20}, A_{20})$ dispersion using the differences of eigenvalues of the dispersion matrix for the time series $A_{-20}(t)$ and $A_{20}(t)$.

Similarity of the distribution of the time series $A_{-20}(t)$ and $A_{20}(t)$ in the sense of asymmetry (skewness) and dispersion around the average value (kurtosis) for group II also regards the depressive configuration of points $p(A_{-20}, A_{20})$ while dissimilarity of the respective time series for group I regards the differences of two tails in points configuration.

Differences of eigenvalues of the dispersion matrix of time series $A_{-20}(t)$ and $A_{20}(t)$ are less in group II (dispersion more regular) than in group I (elongated dispersion). It also regards the analysis of ratio $A_{-20}(t)/A_{20}(t)$ dynamics.

Visual analysis of the influence of the maps scale of potential ranges on the points $p(A_{S1}, A_{S17})$ dispersion confirms that maps generated for EEG segments with of $\pm 30\mu\text{V}$ amplitude should be generated using scale $\pm 15\mu\text{V}$. On the contrary, the change of scale from $\pm 20\mu\text{V}$ to $\pm 40\mu\text{V}$ for the records with maximal amplitude above about $80\mu\text{V}$ decreases the number of points of isopotential areas and causes loss of significant information.

Presented method for analysis of isopotential areas values in sequences of EEG maps based on map histograms allows for evaluation of one of the alternation features, i.e. dynamics of blue and red areas changing. Applying this method also to subareas of EEG maps, e.g. left and right hemispheres or frontal, temporal and occipital regions would allow for topographic estimation of isopotential areas dynamics and therefore for spatio-temporal trends in maps sequences dynamics.

5 Conclusion

Results of this work as well as our previous studies [3, 4] indicate usefulness of isopotential areas analysis for evaluation of the alternation variability in sequences

of the EEG maps. Calculations performed on well selected data [5, 6, 7] enable verification of the effectiveness of the elaborated methods:

- analysis of the plot of ratios of isopotential areas values [3, 4],
- analysis of the scatter plots of isopotential areas values,
- analysis of the influence of maps scale on the visual and quantitative analysis of isopotential areas values.

The results of above examinations suggest application of this method to subareas of individual map or/and to the correlation analysis concerning the localization of isopotential areas. We expect, that it will allow us to identify the configuration of the areas on the maps and their differentiation both quantitative and referring to their topolocalisation.

References

- [1] Duffy, F.H.: *Topographic Mapping of Brain Electric Activity*. Butterworth, Boston (1986)
- [2] Duffy, F.H., Bartels, P.H., Burchfiel, J.L.: Significance probability mapping: an aid in the topographic analysis of brain electrical activity. *Electroencephalogr. Clin. Neurophysiol.* 51(5), 455–462 (1981)
- [3] Goszczynska, H., Doros, M., Kowalczyk, L., Kolebska, K., Dec, S., Zalewska, E., Miszczak, J.: Relationships between Isopotential Areas in EEG Maps before, during and after the Seizure Activity. In: Pietka, E., Kawa, J. (eds.) *Information Technologies in Biomedicine*, pp. 315–324. Springer, Heidelberg (2008)
- [4] Goszczyńska, H., Kowalczyk, L., Doros, M., Kolebska, K., Józwick, A., Dec, S., Zalewska, E., Miszczak, J.: Analysis of variability of isopotential areas features in sequences of EEG maps. In: Kurzynski, M., Wozniak, M. (eds.) *Computer Recognition Systems 3. AISC*, vol. 57, pp. 407–414. Springer, Heidelberg (2009)
- [5] Kowalczyk L.: Analysis of brain bioelectric activity dynamics under stimulation utilizing signal similarity evaluation methods (in Polish). PhD Thesis, Institute of Biocybernetics and Biomedical Engineering Polish Academy of Sciences, Warsaw (2006)
- [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 Discharges In Healthy Subjects. In: 7th European Congress on Epileptology, pp. 47 (s3):48. Helsinki, Epilepsia (2006)
- [7] Kowalczyk, L., Zalewska, E., Miszczak, J., Dec, S.: New criteria for the differentiation between spontaneous EEG in records with seizure activity in healthy subjects and epileptic patients. In: *Proc. of 53 International Congress of Aviation and Space Medicine*, Warsaw, p. 26 (2005)
- [8] Lehmann, D., Pascual-Marqui, R.D., Michel, C.: EEG_microstates. *Scholarpedia* 4(3), 7632 (2009), http://www.scholarpedia.org/article/EEG_microstates
- [9] Lehmann, 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. Springer, Heidelberg (1989)
- [10] 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)

- [11] Lopes da Silva, F.: Computer-Assisted EEG Diagnosis: Pattern Recognition and Brain Mapping. In: Niedermeyer, E., Lopes da Silva, F. (eds.) *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, 5th edn., pp. 1252–1256. Lippincott Williams & Wilkins, Philadelphia (2005)
- [12] Morihisa, J.M., Duffy, F.H., Wyatt, R.J.: Brain electrical activity mapping (BEAM) in schizophrenic patients. *Arch. Gen. Psychiatry* 40(7), 719–728 (1983)
- [13] Murray, M.M., Brunet, D., Michel, C.M.: Topographic ERP Analyses: A Step-by-Step Tutorial Review. *Brain Topogr.* 20, 249–264 (2008)
- [14] Tatsuno, J., Ashida, H., Takao, A.: Objective evaluation of differences in patterns of EEG topographical maps by Mahalanobis distance. *Electroencephalography and Clinical Neurophysiology* 69(3), 287–290 (1988)
- [15] Wackerman, J., Lehmann, D., Michel, C.M., Strik, W.K.: Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. *Int. J. Psychophysiol.* 14, 269–283 (1993)
- [16] Walerjan, P.: First derivation of EEG amplitude - new parameter in EEG mapping (in Polish). *Epileptologia* 8(1), 1–88 (2000)
- [17] Golenia, B., Kosiński, W., Weigl, M.: An Evolutionary Method for Designing Decision Support System. In: Kłopotek, M.A., Przepiórkowski, A., Wierzchoń, S.T., Trojanowski, K. (eds.) *Recent Advances in Intelligent Information Systems*, pp. 275–288. Academic Publishing House, EXIT, Warsaw (2009)

Robust Estimation of Respiratory Rate Based on Linear Regression

J. Łęski, N. Henzel, A. Momot, M. Momot, and J. Śledzik

Abstract. One of the basic parameters which are of very importance in patients monitoring is respiratory rate (frequency of breathing). Previous studies show that there is possibility of estimating the respiratory rate from ECG signal using the modulation amplitude of this signal during the breath. This paper presents the method which is highly sensitive to the ECG amplitude changes and also robust to disturbances, usually accompanying the signal. The idea which distinguishes the method from other is to use changes of ECG amplitude in consecutive heart cycles on the basis of the entire QRS complex. The changes are determined by robust regression analysis (which leads to iterative algorithm) and the regression coefficients are transformed to the frequency domain.

1 Introduction

Respiratory pulse is due to its diagnostic utility, one of the basic parameters to be monitored in patients. Since the other often monitored signal is ECG signal, there are numerous attempts well known in the literature to use this signal to estimate the respiratory rate [8], [23], [6]. The advantage of this approach is the ability to obtain additional information about the patient's condition without the need for additional sensors. Changes in ECG signal (see Fig. 1) caused by breath are induced by thoracic impedance change caused by change in lung volumes during breath and changes of the spatial location of the electrodes against the heart. In addition, there

J. Łęski · N. Henzel

Institute of Electronics, Silesian University of Technology, Poland
e-mail: {jacek.leski, norbert.henzel}@polsl.pl

A. Momot

Institute of Computer Science, Silesian University of Technology, Poland
e-mail: alina.momot@polsl.pl

J. Łęski · N. Henzel · M. Momot · J. Śledzik

Institute of Medical Technology and Equipment, Poland
e-mail: {michal.momot, joanna.sledzik}@itam.zabrze.pl

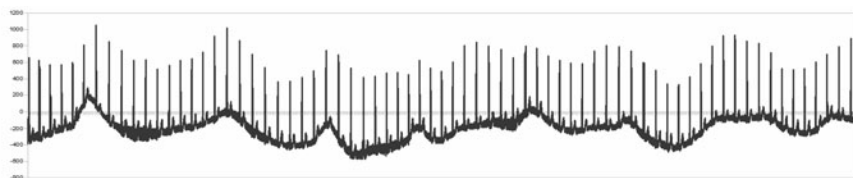


Fig. 1 The ECG recording showing influence of breathing action on R-wave amplitude

is a change in heart rate caused by the influence of the autonomic nervous system. To reconstruct the respiratory rate from ECG signal there is mostly used the modulation amplitude of this signal during the breath (R-wave amplitude or area under the R-wave) [8]. Methods based on changes in heart rate are rarely used due to the possibility of cardiac arrhythmias.

Given the small change of breath induced ECG amplitude modulation and commonly occurring respiratory interference it should be noted that for very few patients analysis of changes in R wave amplitude leads to satisfactory results [8]. The use of the area under the R-waves requires the determining of the characteristic points of the wave (beginning and end of the wave). In the literature there are described algorithms based on the respiratory wave ECG: using wavelet transform [23], adaptive filtering [22], the principal component analysis [13] or the analysis of electric heart vector [2]. The primary drawback of these algorithms is their complexity and large computational effort.

The aim of the study is to introduce the method which is highly sensitive to the ECG amplitude modulation caused by respiratory activity and also robust to interference of this signal. The idea which distinguishes the method from other is to use ECG amplitude changes in consecutive evolutions of heart on the basis of the entire QRS complex using the robust regression function fitting. The performance of developed method was empirically evaluated using the set of artificial as well as real ECG signals.

2 Idea of Proposed Method

The first step of the method is the precise determining of fiducial points for each QRS complex which is realized by maximizing the correlation coefficients between actual QRS complex and the averaged one (see Fig. 2).

After determining fiducial points for each QRS complex, signals are analyzed in time windows for these complexes (in a fixed location with respect to fiducial points, without the need to detect additional characteristic points). Scaling each QRS complex is determined by regression analysis between consecutive QRS complexes (see Fig. 3). Application of the method of least squares is intended to assure robustness of the proposed method to ECG signal interference. This is a basic feature of the method which distinguishes it from the commonly used methods based on determining the QRS amplitude scaling using one sample signal (usually the R wave

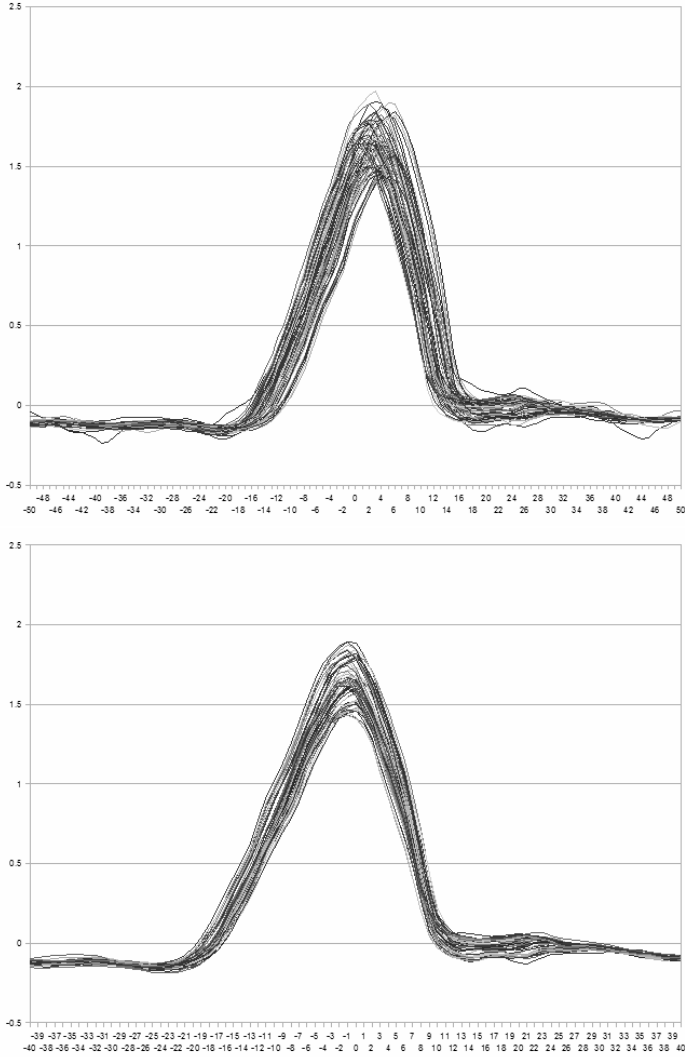


Fig. 2 The alignment of QRS fiducial points (above – inaccurate, below – precise)

amplitude). The primary problem is muscular disturbances and impulsive noise, since other types of interference can be effectively suppressed. The impulsive noise can be an effect of electromagnetic interference caused by activity of electronic devices, such as cellular phones. Assuming that the muscular disturbances can be characterized by additive Gaussian noise, the method of least squares leads to the unbiased minimum variance estimation of ECG amplitude modulating effect caused

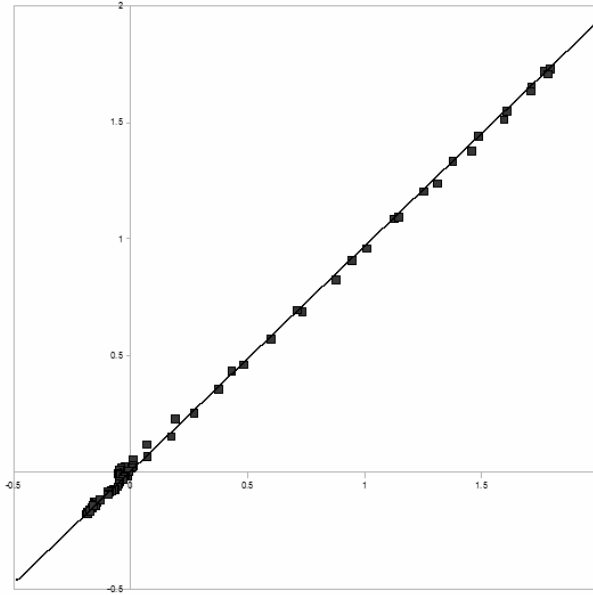


Fig. 3 The example of linear regression between two consecutive QRS complexes

by breathing wave. The impulsive noise can be modeled by heavy-tailed distribution, like Cauchy distribution. This gives the motivation for seeking robust version of regression function fitting method. Such robust algorithm, which is the central point of presented work, will be described in detail in the next section.

The next step of the method is to implement the methods of Fourier analysis on the resulting sequence of QRS scale factors (see Fig. 4). There is determined the maximum absolute value component of the spectral characteristic. If the value exceeds a specified threshold, it is possible to estimate the frequency of respiratory rate. Otherwise, the method does not give valid response.

3 Detailed Description of Algorithm

The first step of presented algorithm is the precise determining of fiducial points for each QRS complex. The performance of this step is highly influenced by effectiveness of baseline wander and powerline interference reduction. The baseline wander is mainly caused by varying with time electrode-skin impedance, breath and patient's movements during signal acquisition [1], [7], [16], [18]. The frequency range of baseline wander is usually less than 1.0 Hz, however, for exercise ECG this range can be wider [1], [12], [21].

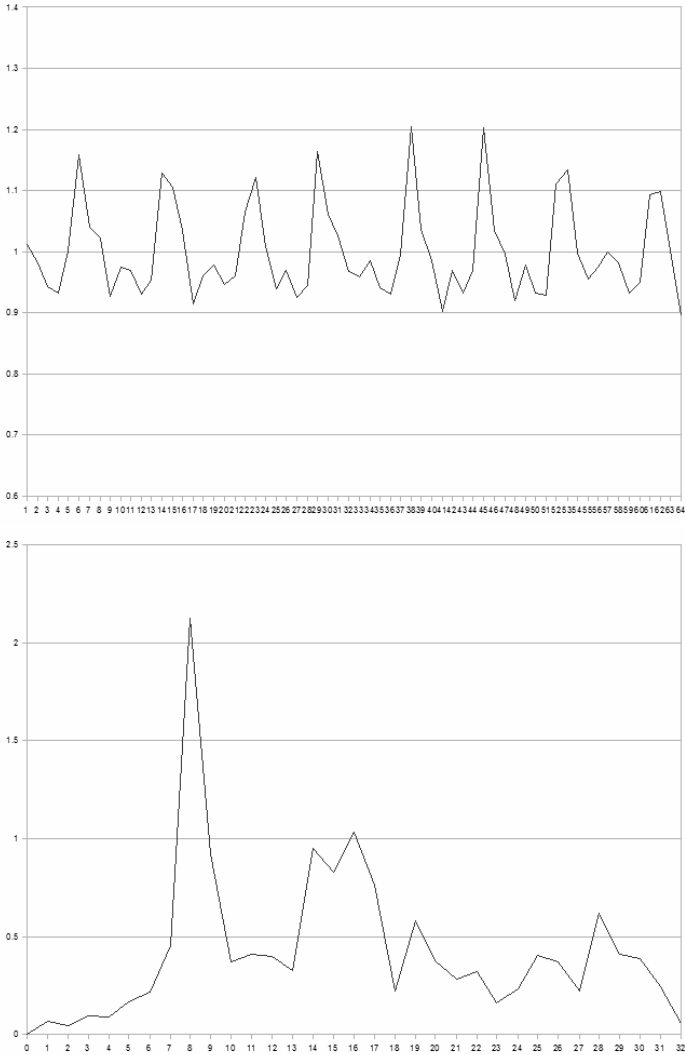


Fig. 4 The series of regression coefficients (above) and its spectrum (below)

The ECG signal is also disturbed by additive 50 or 60 Hz powerline (AC) interference. This kind of disturbance can be modeled by a sinusoid with respective frequency and random phase.

A very simple and efficient method for baseline wander reduction is the high-pass filtering based on a moving average (MA) filter [9], [20], [5].

In this work we used a digital filter with a true linear phase characteristics proposed in [14]:

$$\begin{cases} y_n = 2y_{n-b} - y_{n-2b} + x_n - 2x_{n-a-b} + x_{n-2a-2b}, \\ z_n = x_{n-a} - \frac{y_n}{\left(\frac{a}{b} + 1\right)^2}. \end{cases} \quad (1)$$

where x_n denotes a discrete input signal and z_n is a discrete output signal. Parameters a, b depend on the signal sampling frequency and specify the frequency characteristics of the desired filter. In this work the values for these parameters were, respectively, equal to $a = 120, b = 5$ (according to the sampling frequency of the signal, which was equal to 250Hz).

Detection of QRS complexes is realized by modified version of method described in [19]. Such method involves constructing the detection function using nonlinear function of analyzed signal together with its first two derivatives as well as dynamically adapted threshold. In order to reduce the method's sensitivity to signal disturbances it was developed algorithm of numerical computation of signal derivative based on finite sample set. This algorithm will be described in detail in Section 4.

After QRS complex detection the fiducial points for QRS complexes were determined. The fiducial points were preliminary obtained by the matching filtering method and in the second pass precisely by the Fourier Shift Method [14]. These points localize the signal fragments analyzed on time axis. Thus, we obtain in this way a set of signal cycles denoted as: $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,m}]^\top; i = 1, 2, \dots, N$, where $x_{i,j}$ denotes j th signal sample in i th cycle; N stands for a number of analyzed cycles. In our case, $m = 35$ and a fiducial point for each cycle is located on $j = 18$.

We assume the following linear model of amplitude changes in subsequent cycles

$$\mathbf{x}_i = a_i \mathbf{x}_1 + \boldsymbol{\eta}_i + \mathbf{e}_i; \quad i = 2, 3, \dots, N, \quad (2)$$

where a_i and $\boldsymbol{\eta}_i$ are parameters, which promote amplitude scaling and shifting, respectively. \mathbf{e}_i stands for a random vector with mean equals zero.

Defining the augmented signal vector $\tilde{\mathbf{x}}_i = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{1} \end{bmatrix}$ and the augmented vector of parameters $\tilde{\mathbf{a}}_i = [a_i, \boldsymbol{\eta}_i]^\top$, linear model (2) can be written as

$$\mathbf{x}_i = \tilde{\mathbf{x}}_1 \tilde{\mathbf{a}}_i + \mathbf{e}_i; \quad i = 2, 3, \dots, N. \quad (3)$$

We seek a vector $\tilde{\mathbf{a}}_i$ such that corresponding model fits the data best. If we choose a weighted quadratic loss function then in matrix notation we obtain

$$\min_{\tilde{\mathbf{a}}_i \in \mathbb{R}^2} J(\tilde{\mathbf{a}}_i) \triangleq (\mathbf{x}_i - \tilde{\mathbf{x}}_1 \tilde{\mathbf{a}}_i)^\top \mathbf{G}_i (\mathbf{x}_i - \tilde{\mathbf{x}}_1 \tilde{\mathbf{a}}_i) = \mathbf{e}_i^\top \mathbf{G}_i \mathbf{e}_i, \quad (4)$$

where the matrix $\mathbf{G}_i = \text{diag}(g_{i,1}, g_{i,2}, \dots, g_{i,m})$. Usually, the density function of the residuals \mathbf{e}_i is unknown. To overcome this problem, we need the so-called a robust estimator [10]. Through respective selection of parameters g_i we may to change

different loss functions to the quadratic loss [15]. To obtain robustness to noised data and outliers we use the absolute error function. This loss function is easy to obtain by taking

$$g_{i,j} = \begin{cases} 0, & e_{i,j} = 0, \\ 1/|e_{i,j}|, & e_{i,j} \neq 0. \end{cases} \quad (5)$$

The values of the parameters $g_{i,j}$ depend to residuals of the model. In turn, the residuals depend to $\tilde{\mathbf{a}}_i$. Thus, criterion function (4) should only be minimized by iteratively. The condition for optimality of (4) is obtained by its differentiating with respect to $\tilde{\mathbf{a}}_i$ and setting the result equals to zero

$$\tilde{\mathbf{a}}_i = \left(\tilde{\mathbf{x}}_1^\top \mathbf{G}_i \tilde{\mathbf{x}}_1 \right)^{-1} \tilde{\mathbf{x}}_1^\top \mathbf{G}_i \mathbf{x}_i. \quad (6)$$

The procedure for estimation of $\tilde{\mathbf{a}}_i$ can be summarized in the following steps:

1. Set $\mathbf{G}_i = \mathbf{I}$. Set the iteration index $k = 0$ and $\tilde{\mathbf{a}}_i^{\text{old}} = [0, 0]^\top$.
2. $\tilde{\mathbf{a}}_i = \left(\tilde{\mathbf{x}}_1^\top \mathbf{G}_i \tilde{\mathbf{x}}_1 \right)^{-1} \tilde{\mathbf{x}}_1^\top \mathbf{G}_i \mathbf{x}_i$.
3. $\mathbf{e}_i = \mathbf{x}_i - \tilde{\mathbf{x}}_1 \tilde{\mathbf{a}}_i$.
4. $\mathbf{G}_i = \text{diag}(g_{i,1}, g_{i,2}, \dots, g_{i,m})$, where $g_{i,j}$, for $i = 1, 2, \dots, m$ is obtained by (5).
5. if $k > 1$ and $\left\| \tilde{\mathbf{a}}_i - \tilde{\mathbf{a}}_i^{\text{old}} \right\|_2 < \xi$, then stop
else $\tilde{\mathbf{a}}_i^{\text{old}} = \tilde{\mathbf{a}}_i$ go to (2).

Remarks. The iterations were stopped as soon as the Euclidean norm in a successive pair of $\tilde{\mathbf{a}}_i$ vectors is less than ξ . The quantity ξ is a pre-set small positive value. In all experiments $\xi = 10^{-3}$ is used. The above algorithm requires the inversion of an 2×2 matrix that may be obtained analytically.

The above mentioned algorithm is used for $i = 2, 3, \dots, N$. For $i = 1$ a trivial solution $\tilde{\mathbf{a}}_1 = [1, 0]^\top$ is obtained. Assuming that signal noise has zero-mean the time series $\{1, a_2, \dots, a_N\}$ represents estimation of ECG modulating effect caused by breathing wave.

The final step of the proposed method is to perform a spectral analysis of the obtained sequence of QRS scale factors. It should be noted that this sequence represents a non-uniformly sampled signal. In this case a Fourier analysis is not the best choice, but our experiments confirmed that satisfactory result could be obtained using this approach. A usual discrete Fourier transform was used [17]:

$$R(k) = \sum_{n=0}^{N-1} a_{n+1} \exp(-j2\pi kn/N), \quad (7)$$

where $0 \leq k \leq N-1, 0 \leq n \leq N-1$. The frequency of the respiratory rhythm is determined by the maximum absolute value of spectral component exceeding a specified threshold.

4 Numerical Differentiation Method

The simplest methods of numerical differentiation, such as difference between two consecutive samples, are usually highly sensitive to the presence of signal disturbances, especially high-frequency noise. To overcome this problem the more sophisticated algorithms [3], [4], including differential quadratures based on polynomial interpolation or approximation [24] are incorporated.

The proposed signal differentiation method exploits Lagrange polynomial interpolation. The nodes of interpolation are equidistant within symmetric fixed-width window. The first derivative of such polynomial can be calculated analytically and it is taken as an estimate of signal derivative. The algorithm of computing this value is briefly described below.

Let D denote the fixed radius of time window around the N th sample of signal. The values of interpolated function within this window constitute a vector $\mathbf{y} = [y_{N-D}, \dots, y_N, \dots, y_{N+D}]$, the degree of polynomial g is determined by number of samples and it is equal to $2D$, therefore $g(x) = \sum_{j=0}^{2D} a_j x^j$ and it is equivalent to vector $[a_0, a_1, \dots, a_{2D-1}, a_{2D}]$. The nodes of interpolation are equidistant and each node is equivalent to a single sample index. The assumption could be made, without any loss of generality, that nodes are of the form:

$$x_{N-D} = -D, \dots, x_{N-1} = -1, x_N = 0, x_{N+1} = 1, \dots, x_{N+D} = D. \quad (8)$$

This makes the interpolation method time-invariant and guarantees that only the values of interpolated function in fixed-radius neighborhood will be taken into account. Vector of interpolation polynomial coefficients can be easily obtained by solving the system of linear equations $\mathbf{V}\mathbf{a} = \mathbf{y}$, where $(\mathbf{V})_{i,j} = i^j$ for $i = -R, \dots, R$, $j = 0, \dots, 2R$, i.e.

$$\mathbf{V} = \begin{bmatrix} -1 & -D & \dots & (-D)^{2D-1} & (-D)^{2D} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 1 & \dots & -1 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & D & \dots & D^{2D-1} & D^{2D} \end{bmatrix}. \quad (9)$$

The matrix \mathbf{V} might be easily proven to be non-singular and this leads to solution in the form $\mathbf{a} = \mathbf{V}^{-1}\mathbf{y}$. Moreover, the primary objective of this procedure, which is the estimating of function derivative in a center of time window, does not necessarily require determining all polynomial coefficients explicitly, since

$$g'(0) = a_1 = \mathbf{d}^T \mathbf{V}^{-1} \mathbf{y}, \quad (10)$$

where $\mathbf{d}^T = [0, 1, 0, \dots, 0]$. The numerical differentiation algorithm could be seen as high-pass filtering procedure using FIR filter with constant coefficients. These coefficients form the vector $\mathbf{d}^T \mathbf{V}^{-1}$. To estimate the second (or even higher order) derivative, the above described procedure can be repeated iteratively.

5 Results of Experiments

There were conducted several experiments for artificial ECG signals (of known frequency modulation amplitude) and the real ECG signals. In the latter case, the patient's breathing was synchronized with the metronome. In the case of artificial signal as the ECG signal there was taken ANE20000 from database CTS [11]. There were created 50 time series, each of them containing 60 consecutive identical copies with added artificial noise (being the random variables generated using the mixture of Gaussian and Cauchy probability distribution) and artificial respiratory wave. The average relative error was 7.4%. For the real ECG signals there were taken 53 recordings, each of them lasting over one minute. In this case the average relative error was 11.9%.

6 Conclusions

Conducted experiments give reason to hope that the proposed method of assessment of respiratory frequency can be used in devices to monitor the patient's ECG signal, where no additional sensors are required. Although the presented method uses the information contained in single channel ECG signal, it is possible to develop a modified version of the method for multiple leads signal which may lead to improving the method performance. It also seems that the performance of the proposed method can be enhanced in by modifying the methods of robust regression function fitting. This will be the subject of further studies and evaluation.

Acknowledgements. This research was partially supported by Polish Ministry of Science and Higher Education as Research Project N N518 406438.

References

- [1] van Alste, J.A., van Eck, W., Herrmann, O.E.: ECG Baseline Wander Reduction Using Linear Phase Filters. *Compt. Biomed. Res.* 19, 417–427 (1986)
- [2] Bailón, R., Sornmo, L., Laguna, P.: A Robust Method for ECG-Based Estimation of the Respiratory Frequency During Stress Testing. *IEEE Trans. Bio. Eng.* 53(7), 1273–1285 (2006)
- [3] Braci, M., Diop, S.: On numerical differentiation algorithms for nonlinear estimation. In: *Proc. 42nd IEEE Conf. on Decision and Control, Maui, Hawaii, USA*, pp. 2896–2901 (2003)
- [4] Burden, R.L., Faires, J.D.: *Numerical Analysis*. Brooks/Cole, Monterey (2000)

- [5] Canan, S., Ozbay, Y., Karlik, B.: A Method for Removing Low Frequency Trend From ECG Signal. In: Proc. of Int. Conf. Biomed. Engin. Days, pp.144–146 (1998)
- [6] Chazal, P., Heneghan, C., Sheridan, E., Reilly, R., Nolan, P., O'Malley, M.: Automated processing of single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Trans. Bio. Eng.* 50(6), 686–696 (2003)
- [7] Ciarlini, P., Barone, P.: A Recursive Algorithm to Compute the Baseline Drift in Recorded Biological Signals. *Compt. Biomed. Res.* 21, 221–226 (1988)
- [8] Correa, L.S., Laciari, E.T., Jane, R.: Performance evaluation of three methods for respiratory signal estimation from the electrocardiogram. In: 30th Ann. Int. Conf. IEEE/EMBS, pp. 4760–4763 (2008)
- [9] Frankiewicz, Z.: Methods for ECG signal analysis in the presence of noise. Ph.D. Thesis, Silesian Technical University, Gliwice (1987)
- [10] Huber, P.J.: *Robust statistics*. Wiley, New York (1981)
- [11] International Electrotechnical Commission Standard 60601-3-2 (1999)
- [12] Laguna, P., Jane, R., Caminal, P.: Adaptive Filtering of ECG Baseline Wander. In: Proc. of Int. Conf. of the IEEE Eng. in Med. and Biol. Soc., pp. 508–509 (1992)
- [13] Langley, P., Bowers, E.J., Murray, A.: Principal component analysis as a tool for analysing beat-to-beat changes in electrocardiogram features: application to electrocardiogram derived respiration. *IEEE Trans. Bio. Eng.* 57(4), 821–829 (2010)
- [14] Leski, J.M.: A new possibility of non-invasive electrocardiological diagnosis. Report no.1233, Silesian University of Technology, Gliwice (1994)
- [15] Leski, J.M., Henzel, N.: Generalized ordered linear regression with regularization. Submitted to *IEEE Trans. Syst., Man and Cybern.* (2010)
- [16] Meyer, C.R., Keiser, H.N.: Electrocardiogram Baseline Noise Estimation and Removal Using Cubic Splines and State-Space Computation Techniques. *Compt. Biomed. Res.* 10, 459–470 (1977)
- [17] Mitra, S.K., Kaiser, J.F. (eds.): *Handbook for digital signal processing*. John Wiley & Sons, New York (1993)
- [18] Oguz, S.H., Asyali, M.H.: A Morphology Based Algorithm for Baseline Wander Elimination in ECG Records. In: Proc. of Int. Conf. Biomed. Engin. Days, pp. 157–160 (1992)
- [19] Pan, J., Tompkins, W.J.: A Real-Time QRS Detection Algorithm. *IEEE Trans. Biomed. Eng.* 32(3), 230–236 (1985)
- [20] Riccio, M.L., Belina, J.C.: A Versatile Design method of Fast, Linear-Phase FIR Filtering Systems for Electrocardiogram Acquisition and Analysis Systems. In: Proc. of Int. Conf. Comput. Cardiol., pp. 147–150 (1992)
- [21] Thakor, N.V., Zhu, Y.S.: Application of Adaptive Filtering to ECG Analysis: Noise Cancellation and Arrhythmia Detection. *IEEE Trans. Biomed. Eng.* 38(8), 785–794 (1991)
- [22] Varanini, M., Emdin, M., Allegri, F., Raciti, M., Conforti, F., Macerata, A., Taddei, A., Francesconi, R., Kraft, G., Abbate, A.L., Marchesi, C.: Adaptive filtering of ECG signal for deriving respiratory activity. *Comp. Cardiol.*, 621–624 (1990)
- [23] Yi, W.J., Park, K.S.: Derivation of respiration from ECG measured without subject apos's awareness using wavelet transform. In: 24th Ann. Conf. EMBS/BMES, pp. 130–131 (2002)
- [24] Zhang, Y.: *Advanced Differential Quadrature Methods*. CRC Press, Boca Raton (2009)

Discriminating Power of an Sample Entropy and a Nonlinear Association Index in Prediction of a Preterm Labor Based on Electrohysterographical Signals

Dariusz Radomski and Agnieszka Małkiewicz

Abstract. Recognition of physiological and pathological patterns in biomedical signals is still a challenge. There are many linear and nonlinear techniques proposed for this purpose but their effectiveness is seldom compared and ranked. The aim of the paper was to compare a discriminating power of a sample entropy and a nonlinear association index in prediction of a labor based on electrohysterographical (EHG) signals. The EHG signals were registered at women being during a labor or waiting for beginning of a labor. A sample entropy was estimated for a single component of an EHG signal. A nonlinear association index was computed to express a plausible relation between two components of an EHG signal. The comparison of usefulness of these parameters in a labor prediction was performed using ROC. The obtained results show that a labor prediction based on the nonlinear association index is more effective than using the sample entropy.

1 Introduction

The main area of application of pattern recognition techniques is medicine. These tools are usually used in a clinical diagnostic process or in prognosis of a patient's health status being treated for a given disease. Pattern recognition methods applied in a diagnostic process play several roles. They are helpful for finding a pattern of diagnostic symptoms which is specified for a given disease creating a diagnostic model or guidelines for clinical protocols. Having an identified of diagnostic models for different diseases, these techniques are used as a classification tool to recognize the most probable disease in a given patient. Moreover, the more advanced methods aid to extract clinically useful information from biological signals registered with a patient.

Pattern recognition systems are most often used in several medical disciplines such as radiology, cardiology, neurology, anesthesiology or oncology [1]. However, they

Dariusz Radomski · Agnieszka Małkiewicz
Institute of Radioelectronics Warsaw University of Technology
Nowowiejska 15/19 00-665 Warsaw, Poland
e-mail: D.Radomski@ire.pw.edu.pl

are rarely applied in obstetrics and gynecology. It can be presumed that this disproportion arises from historical and cultural conditions because human reproduction was always treated as a magic phenomenon too private for science investigations.

Generally, a pattern recognition can be an efficient clinical tool for monitoring of two biological components: health of a fetus and health of a mother with particular attention to her reproductive system.

Evaluation of anatomical and physiological states a mother reproductive system is performed basing on USG images and measuring of uterine activity. Qualitative and quantitative knowledge of uterine contraction is necessary to predict a course of term labor or to predict a risk of a preterm labor.

Routinely, uterine contractions are monitored by a mechanical sensor placed on an abdominal skin of a pregnant woman. This sensor measures an uterine wall deformation. Unfortunately, the sensitivity of this method is low particularly for obese women.

Thus, an alternative method for monitoring of uterine contractions is developing. Presuming that each uterine contraction is preceded by bioelectrical activity of an uterine muscle (called myometrium) we can measure these bioelectrical impulses instead of mechanical deformations. This method is called electrohysterography (EHG) and it is similar to well known electrocardiography. However, in opposite to ECG there is no repeatable signal pattern which is uniquely associated with uterine contractions. Thus, the main goal of EHG pattern recognition is to find such parameterization of EHG signals which enables to predict an upcoming labor activity of a pregnant uterus.

Up to the present, there were proposed many ways of EHG parameterization basing on Fourier or wavelet spectra of the signal as well as some nonlinear techniques [2], [3]. For example, recently, Radomski *et al.* suggested an application of sample entropy index or nonlinear association index could be used as parameters differentiating an activity of a pregnant uterus [4], [5]. Despite of promising results of many published methods there is still unknown which EHG parameterization enables for the best prognosis of an upcoming labor.

The aim of the work was to compare a discriminating power of an upcoming labor prediction based on two nonlinear methods of EHG parameterization. The first one was based on information gathered in a single EHG signal which was represented by an sample entropy index. The second used information about a nonlinear association between two components of EHG signals. The comparison of discriminating power was performed with help of areas under receiver – operating curves.

2 Methods

2.1 Sample Entropy Estimation

The idea supporting application of sample entropy to EHG parameterization was presented in [5]. The sample entropy was estimated in similar way as an approximated entropy [3]. The estimation procedure was conducted in the following manner.

Let a single component of a EHG signal be represented by a time series denoting as $\{ehg(n)\}$. Let's create m vectors contained consecutive values of ehg_i , commencing at the i -th sample, i.e.

$$\mathbf{EHG}_m(i) = \left[ehg(i) \ ehg(i+1) \ \dots \ ehg(i+m-1) \right] \tag{1}$$

for $1 \leq i \leq N - m + 1$

By $d[\mathbf{EHG}_m(i), \mathbf{EHG}_m(j)]$ is denoted the distance between two vectors $\mathbf{EHG}_m(i)$, $\mathbf{EHG}_m(j)$ which is defined as:

$$d[\mathbf{EHG}_m(i), \mathbf{EHG}_m(j)] = \max_k |ehg(i+k) - ehg(j+k)| \tag{2}$$

The distance measure was used for counting of the number of the similar elements of the vectors $\mathbf{EHG}_m(i)$ and $\mathbf{EHG}_m(j)$.

Let introduce a set of such indexes which for this distance is not greater than r . This set is described by the following expression:

$$J_m = \{j : d[\mathbf{EHG}_m(i), \mathbf{EHG}_m(j)] \leq r\} \tag{3}$$

For a given \mathbf{EHG}_m and for $1 \leq i \neq j \leq N - m$ we can define the coefficient

$$B_i^m = \frac{\text{card}\{J_m\}}{N - m - 1}. \tag{4}$$

Then, one can compute the number of the similar vector elements averaging over i in the following manner:

$$B^m = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m \tag{5}$$

It expresses the probability that two sequences coincide for m points. Analogically, such probability is computed for the vector \mathbf{EHG}_{m+1} . The estimator of the sample entropy is given by;

$$\text{SamEn}(m, r) = -\ln \frac{B^{m+1}}{B^m} \tag{6}$$

This estimator depends on two parameters m, r . The results of sensitivity analysis obtained by Radomski showed the following best values of these parameters $m = 2, r = 0.17\sigma$, where σ is a variance of the parameterized EHG signal.

2.2 Estimation of a Nonlinear Association Index Estimation

The estimation method of a nonlinear association index was firstly proposed by Kalitzin *et al.* and was used to analyze relations between EEG components [7]. This index is denoted by h^2 .

Let $ehg_1(n) : n = 1, \dots, N$ denotes a value of EHG signal measured at the n^{th} moment at an uterine fundus. Similarly, $ehg_2(n) : n = 1, \dots, N$ denotes a value of EHG signal measured at the n^{th} moment at an uterine cervix. Then, the h^2 index is computed in the following manner:

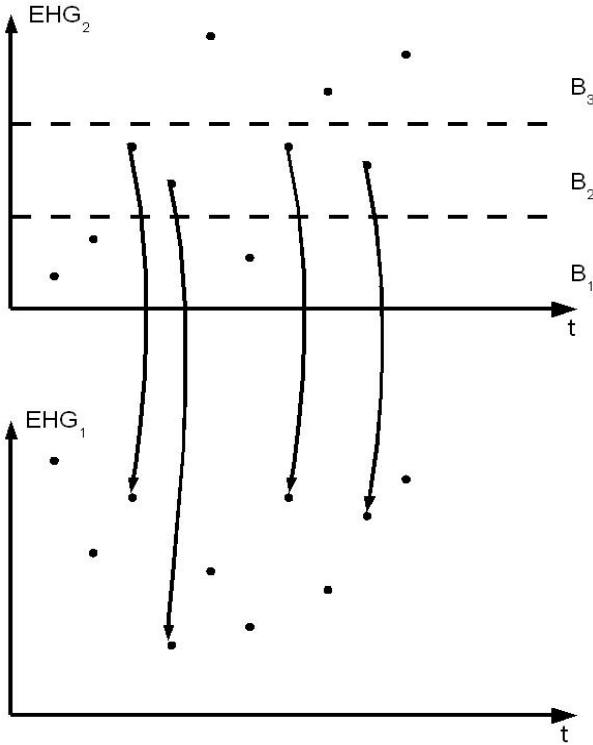


Fig. 1 The scheme of EHG samples selection used for calculation of h^2 index

$$h^2(EHG_1|EHG_2) = 1 - \frac{\frac{1}{N} \sum_{a=1}^M \left[\sum_{n: EHG_2 \in B_a} (EHG_1(n) - \langle EHG_1 \rangle_a)^2 \right]}{\sigma_{EHG_1}^2}$$

$$\langle EHG_1 \rangle_a = \frac{\sum_{n: EHG_1 \in B_a} EHG_1(n)}{N_a}, \quad \sum_a N_a = N \quad (7)$$

Fig. 1 presents the scheme for selection of EHG_1 samples which correspond to the samples of EHG_2 belonging to B_a bin. $\sigma_{EHG_1}^2$ represents a standard deviation of signal EHG_1 .

This index has the following interesting property. It can be proven that when two EHG components are linear correlated then the $h^2(EHG_1|EHG_2) \equiv h^2(EHG_2|EHG_1) \equiv 1$, [7]. Thus, an observation that

$$h^2(EHG_1|EHG_2) \neq h^2(EHG_2|EHG_1) \quad (8)$$

is a circumstance for identification of a nonlinear relationship between the analyzed components.

Estimating h^2 index for various time delays between two components of EHG signal we obtained nonlinear cross-correlation sequences. The maximal value of the nonlinear cross-correlation sequences was used as a pattern of the 2D EHG signal.

2.3 EHG Signal Acquisition

Acquisition of EHG signal was made using the specialized device designed by the Institute of Medical Technology and Equipment ITAM [6]. Biopotentials generated by myometrium were registered by two Ag-Cl electrodes placed on an abdominal skin. The first electrode was placed over an uterine fundus and the other one over an uterine cervix. Because EHG signals were measured differently the reference electrode was placed on a patient's thigh. Mechanical uterine activities was also monitored by the TOCO sensor (Fig 2).

Hardware details are described in [7]. The measured biopotentials were amplified ($k_u = 2400 [V/V]$) and filtered by a lowpass filter (0-5Hz). The antialiasing filter was applied too. The sampling frequency was 10 Hz.

The EHG signals were registered in two groups of patients. The first group contains women waiting in a hospital for a term labor. The other group consisted of EHG signal registered during 2nd period of a physiological labor.

The Fig 3 presents an example of mechanical and electrical uterine activity registered during 2nd period of a physiological labor. We can observe a mechanical contraction is associated with increased bioelectrical spikes.

Electrohysterographical signals were registered in a delivery room at university clinics of gynecology and obstetrics. All examined patients were healthy and they had unifetal pregnancies. Registration of EHG signals was performed in two groups of patients. In the first group of 15 women EHG measurement was made during 2nd period of a labor. The control group consisted of 29 pregnant women expecting the beginning of a labor.

3 Results

A sample entropy was estimated for the single EHG component registered over an uterine fundus because at the beginning of a labor an activity of this uterine part is stronger than an activity of an uterine cervix. This observation supports slightly greater AUC under ROC computed for $h^2 (EHG_1|EHG_2)$ than AUC computed for $h^2 (EHG_2|EHG_2)$.

The comparison of the mean values of the sample entropy and the nonlinear association indexes between two group the signals is presented in the tab 1. The used Mann – Whitney test showed that all differences were statistically significant.

The Fig 4 shows the ROC computed for the both nonlinear association indexes.

Although there are no significant differences between the area under all curves ($p=0.23$) the greatest area is for the nonlinear association index (Tab 2).

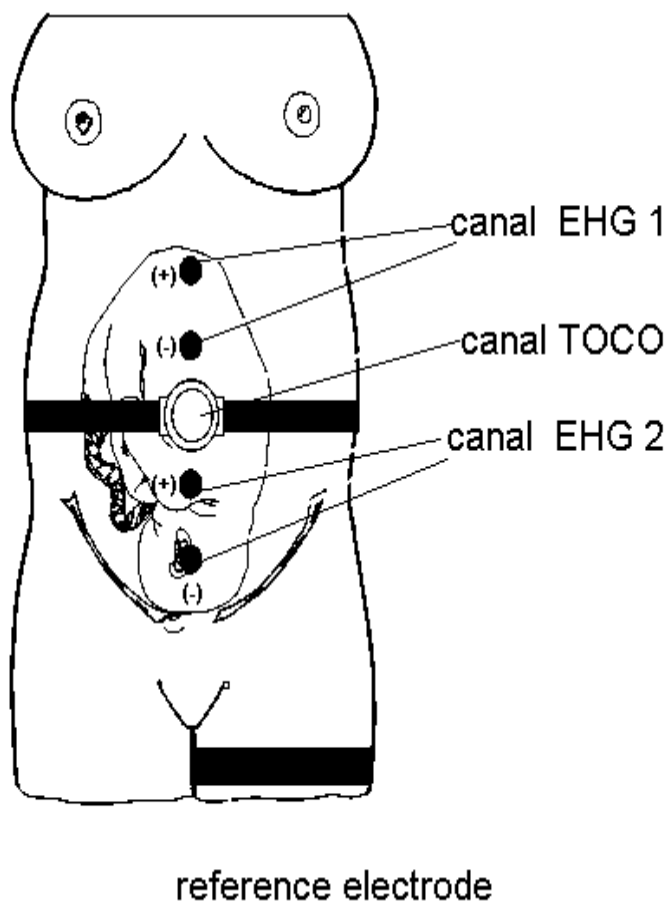


Fig. 2 Location of electrodes during EHG registration

Table 1 The comparison of the parameters values computed before or during a labor

	Mean labor expected	Mean during labor expected	Std .dev. during expected	Std .dev. during labor	p-value
$h^2 (EHG_2 EHG_1)$	0,087	0,323	0,115	0,277	0,000231
$h^2 (EHG_1 EHG_2)$	0,087	0,349	0,136	0,301	0,000168
Sample entropy	1,083	0,744	0,453	0,480	0,008678

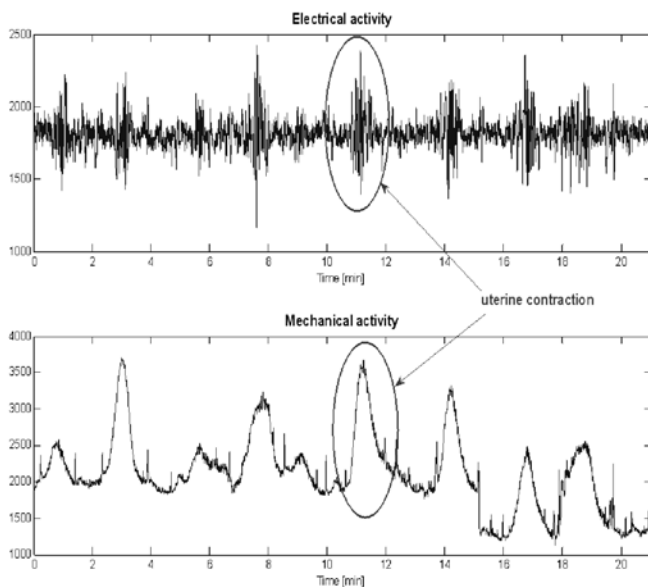


Fig. 3 An example of electrical and mechanical uterine activity during a labor

Table 2 The AUC computed for the studied parameters

Parameter	AUC	AUV	std. dev.
$h^2(EHG_2 EHG_1)$	0,83	0,06	
$h^2(EHG_1 EHG_2)$	0,85	0,06	
Sample entropy	0,74	0,08	

Table 3 Quality of a labor recognition based on the cut-off values

Parameter	Sensitivity	Specificity	PPV	NPV
$h^2(EHG_1 EHG_2) = 0,108$	0,800	0,828	0,706	0,889
Sample entropy = 0,771	0,733	0,793	0,647	0,852

The cut-off values were calculated for $h^2(EHG_1|EHG_2)$ and sample entropy. The errors for a labor recognition based on these cut-off values were shown in the Tab.3

These results indicate that a labor prediction on the base of the nonlinear association index is more precise than based on the sample entropy.

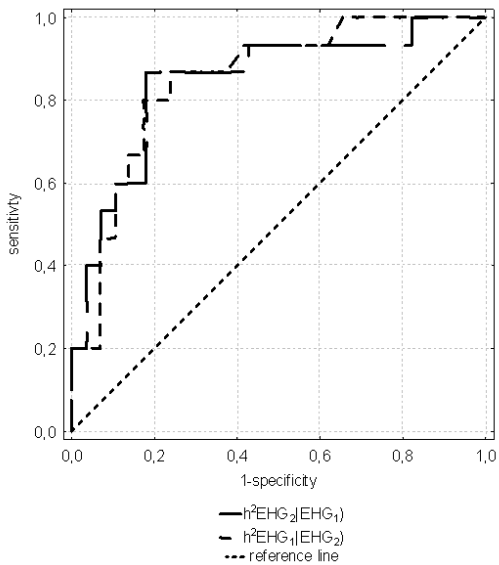


Fig. 4 ROCs for computed for the nonlinear association indexes

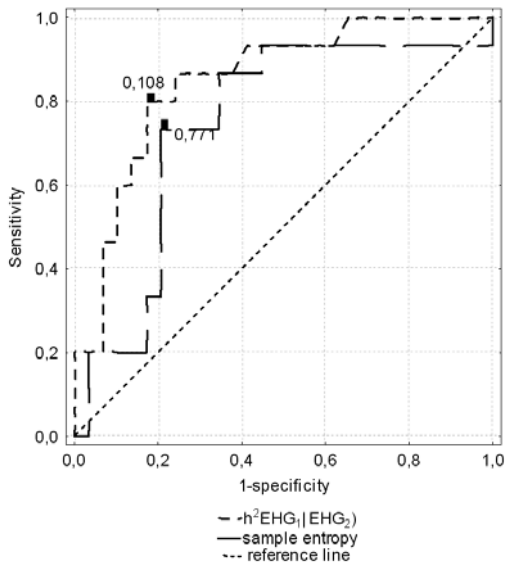


Fig. 5 ROCs computed for sample entropy and for $h^2(EHG_1|EHG_2)$

4 Conclusion

Recognition of physiological and pathological patterns in biological signals is still a serious challenge. Some of biological signals such as ECG have well defined patterns which correspond to a physiological activity of a given organ. Unfortunately, in the most of biological signals such complexes are unknown yet. In these situations authors propose many methods for parameterization of these signals as well as techniques for signal classifications based on these parameters. However, there is perceptible lack of comparative studies showing superiority of one method over others.

The presented work was concentrated on two nonlinear parameterization methods applied to an electrohysterographical signal. We performed comparison of discriminating power of a sample entropy and a nonlinear association index in prediction of a preterm labor based on electrohysterographical signals using the same database of the collected signals.

Although these both methods show sufficient discriminating power the obtained results indicate that prediction based on the proposed nonlinear associated index is better. It is supported by the physiological theory stating that mutual synchronization of a bioelectrical activity of an uterus is necessary to a labor progression [8].

Acknowledgments. This work was supported by the MNiSW grant No. N N518 505339.

References

- [1] Leistriz, L., Galicki, M., Kochs, E., Zwick, E.B., Fitzek, C., Reichenbach, J.R., Witte, H.: Application of generalized dynamic neural networks to biomedical data. *IEEE Trans. Biomed Eng.* 53(11), 2289–2299 (2006)
- [2] Rossi, J., Gondry, J., Baaklini, N., Naepels, P., Marque, C.: Wavelet analysis of electrohysterography of women exhibiting clinical signs of high-risk pregnancy. *Proceedings of the Engineering in Medicine and Biology Society* 2, 1059–1060 (1995)
- [3] Graczyk, S., Jezewski, J., Horoba, K.: Analysis of abdominal electrical activity of uterus—approximate entropy approach. In: *Proceedings of the 22nd Annual International*
- [4] Radomski, D.: Sensitivity Analysis of a Sample Entropy Estimator on its Parameters in Application to Electrohysterographical Signals. *Biocybernetics and Biomedical Engineering*
- [5] Radomski, D., Mańćkiewicz, A.: Identification of a Nonlinear Association Between Components of the Electrohysterographical Signal. *Int. J. Electron. Telecom.* 56(3), 287–290 (2010)
- [6] Zietek, J., Sikora, J., Horoba, K., Matonia, A., Jezewski, J., Magnucki, J., Kobielska, L.: Prognostic value of chosen parameters of mechanical and bioelectrical uterine activity in prediction of threatening preterm labour. *Gin Pol.* 3, 193–200 (2009)
- [7] Kalitzin, S.N., Parra, J., Velis, D.N., Lopes da Silva, F.H.: Quantification of Unidirectional Nonlinear Associations Between Multidimensional Signals. *IEEE Transaction on Biomedical Engineering* 54(3), 454–461 (2007)
- [8] Chard, T., Grudzinskas, J.G.: *The Uterus*. Cambridge University Press, Cambridge (1994)

Knowledge Extraction for Heart Image Segmentation

Arkadiusz Tomczyk and Piotr S. Szczepaniak

Abstract. This paper focuses on the problem of knowledge extraction, which is necessary to find heart ventricles in computed tomography (CT) images. In the proposed approach *potential active contours* are used as a segmentation technique. An energy function used during the evolution of contour requires a proper identification of blood inside heart ventricles as well as an approximate localization of interventricular septum. The methodology effectively allowing to extract that semantic information from the images is described.

1 Introduction

The *active contour* term was proposed by Kass, Witkin and Terzopoulos in [3] to distinguish a new group of image segmentation methods that are able to use high-level knowledge during low-level image segmentation. The basic concept of techniques belonging to that group can be defined as follows: for the image that is segmented define a set of acceptable contours as well as an energy function evaluating those contours and next find the best contour according to that function. In this definition contour is understood as a structure capable of identifying which part of the image should be considered as an object and which should be treated as a background. The simplest contour model used in [3] is a polygon where the interior of the polygon in the image indicates the localization of the sought object. Other models can be found in [1, 2]. In this paper potential contour proposed in [5] is used for this purpose. The choice of contour model is important as it determines the set of objects that can be found. However, the crucial element of the *active contour* methods is a choice of energy function. This function expresses our expectations concerning the searched optimal contour and consequently should contain any available knowledge from the considered domain. This work focuses on the

Arkadiusz Tomczyk · Piotr S. Szczepaniak

Institute of Information Technology, Technical University of Lodz,
Wolczanska 215, 90-924 Lodz, Poland

e-mail: arkadiusz.tomczyk@p.lodz.pl

problem of proper knowledge extraction that enables obtaining satisfying segmentation results of heart ventricle in computed tomography (CT) images using *active contour* approach.

2 Problem

Pulmonary embolism determines a serious diagnostic problem since it is one of the most frequent cause of death among the population in the most developed societies. Development of computed tomography (CT) allowed to introduce new diagnostic methods. One of them bases on the assessment of right ventricle systolic pressure which leads to shifts in interventricular septum curvature. However, that approach requires drawing of endocardial and epicardial contours, which, when it is done manually, is time-consuming as well as prone to errors due to technical mistakes and tiredness. That is why this paper focuses on the task of automating that process. The analyzed image data were obtained using ECG-gated computed tomography scanner in a standard chest protocol after intravenous injection of contrast media. Next, heart cycle was reconstructed in 10 phases and two chamber short axis view is generated leading to 4D image sequence, i.e. the set of 2D images for each reconstructed slice and for each phase of heart cycle. Further image in each phase is called a frame.

The experts knowledge needed to find the proper contours can be expressed using the following statements:

- Contour describing the interior of the ventricle should contain all the pixels representing blood inside that ventricle.
- Contour should be possibly small but smooth since the interior of the ventricle can contain not only the blood but also fragments of heart muscle.
- The blood with injected contrast is represented by bright pixels.
- The interventricular septum is a part of heart muscle separating both ventricles.
- The interventricular is represented by darker pixels between the interiors of left and right ventricle.

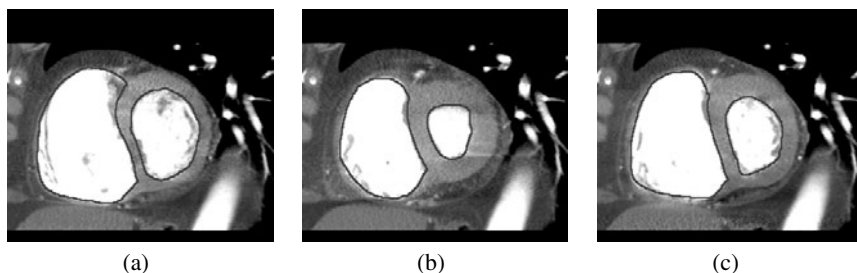


Fig. 1 Sample images of the same heart (the same slice and three different phases of heart cycle) with contours drawn by an expert.

In other words the searched contour should be smooth, small and contain all those bright pixels inside the segmented ventricle that lie on the proper side of interventricular septum. Sample contours drawn by an expert for three different frames of the same slice are presented in Fig. 1.

Authors of this work have already proposed a simple solution of that problem in [6]. In that paper blood with contrast was detected using threshold defined by an expert for each 4D sequence. That makes the process of segmentation semi-automatic but selecting the threshold is of course less tiresome than drawing all the contours. In this work the similar approach is used. Images after thresholding are depicted in Fig. 2. In [6] those images were used without any further processing and the interventricular septum was approximated using straight line. In this work both those steps are replaced by more sophisticated methods of knowledge extraction.

3 Ventricle Interior

The analysis of the images after thresholding reveals that this simple process does not allow to detect only blood with contrast inside ventricles but also other structures inside the body. For example in Fig. 2 these are: ribs on the left side of the image, blood vessels in left lung on the right side and organs in abdominal cavity at the bottom. To remove those structures two-phase preprocessing algorithm was used. Both phases utilize the fact that all the images are a part of 4D image sequence.

In the first phase it was noticed that for the same slice during the heart cycle the only changing part of the images are the regions representing heart. This observation allows to identify those parts of the images that are moving. To find them for each slice those pixels are sought that do not change on 9 from 10 frames. These pixels are depicted in Fig. 3 with light and dark gray color. It is evident that this procedure will not find only not moving fragments outside the heart (dark gray color) but it will find also some regions inside the heart (light gray color). To distinguish them it was observed that the not changing regions inside heart are surrounded by adjacent, relatively large (in this work regions with area greater than 200 pixels were considered), changing regions (black color). Those regions were detected using region



Fig. 2 Sample images after thresholding.

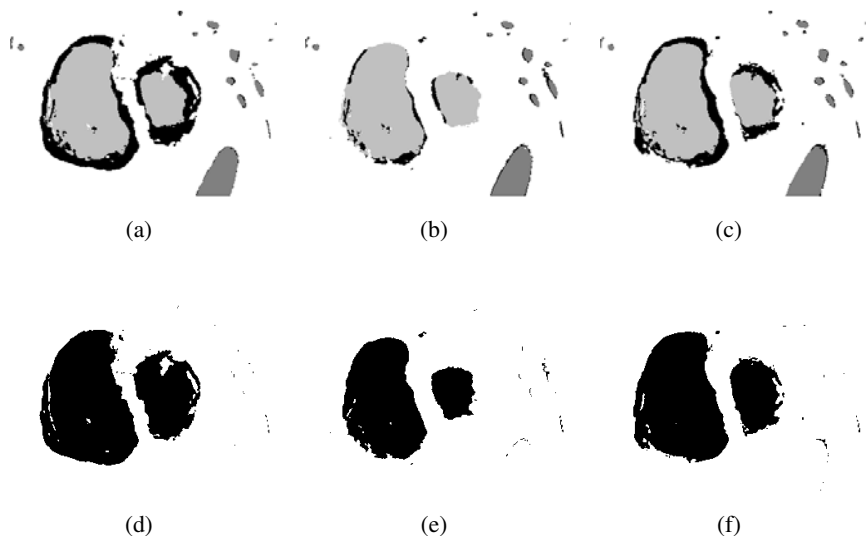


Fig. 3 First phase of preprocessing: (a), (b), (c) - moving and not moving regions, (d), (e), (f) - results of preprocessing.

growing algorithm. The results after removing not moving regions outside the heart are also presented in Fig. 3.

The first phase was not able to remove all the undesirable regions. It left small artifacts that either were not considered as a stationary regions or represent small regions surrounding removed stationary regions. Their existence can be explained by small movements of patient body during the examination or by inaccuracies in reconstruction process. To remove those artifacts the second preprocessing phase was applied. It could not remove small regions because small regions could be also a part of heart ventricle. To avoid that problem first for each slice the sum of 10 frames after first phase was calculated. This sum is presented in Fig. 4 where pixels being a part of that sum and not present in the given image are marked with light gray color. Next, one or two biggest regions (here only regions with area greater than 2000 pixels) were detected using region growing algorithm and it was assumed that they contain ventricles inside. Other regions lying relatively far (in the distance greater than 5 pixels) were removed (dark gray pixels). This procedure allowed to achieve results presented also in Fig. 4.

To evaluate the presented approach it was applied for 15 image sequences with 80 images in each sequence which gives 1200 images. For each image it was checked how many pixels, marked by an expert as pixels lying inside of the ventricles, was removed. The average number of such pixels was 5 from 89600 pixels in the image. Distribution of the incorrectly removed pixels is presented in Table 1.

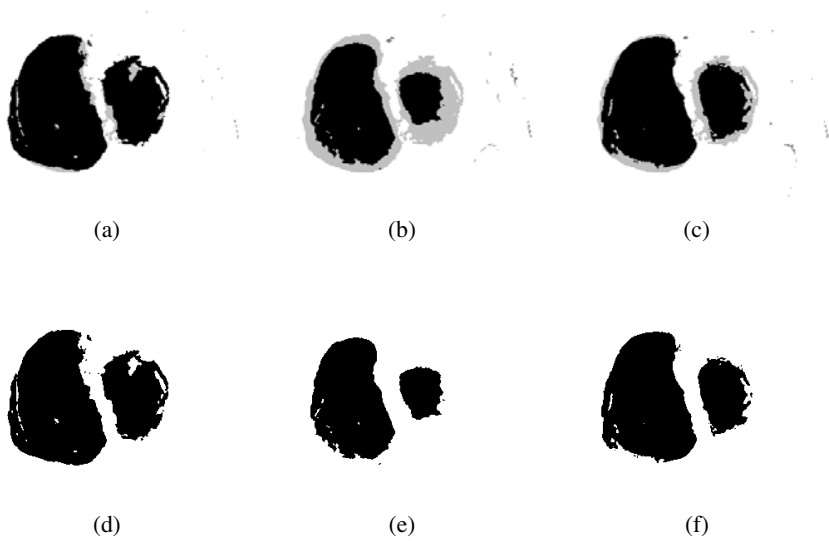


Fig. 4 Second phase of preprocessing: (a), (b), (c) - regions after first phase with pixels being a part of calculated sum for the considered slice, (d), (e), (f) - results of preprocessing.

Table 1 Distribution of incorrectly removed pixels among 1200 analyzed images during the identification of blood inside heart ventricles.

	Number of pixels				
	[0, 20)	[20, 40)	[40, 60)	[60, 80)	[80, ∞)
Number of images	1119	58	12	7	4

4 Interventricular Septum

Having found the blood with contrast inside heart ventricles the next element of knowledge needed during segmentation process is localization of interventricular septum. In this work its localization was approximated using parabola and the search of optimal parabola was again divided into two phases.

In the first phase, using *potential active contour* approach introduced in [5], the smallest circular contour containing both ventricles was sought. The circular shape of contour was obtained as only one object source and one background source was considered. The energy function in this case was defined as:

$$E(c) = wE^{out}(c) + E^{area}(c) \tag{1}$$

where E^{out} component calculated how many pixels representing blood inside ventricles lie outside the contour and E^{area} component calculated the number of pixels

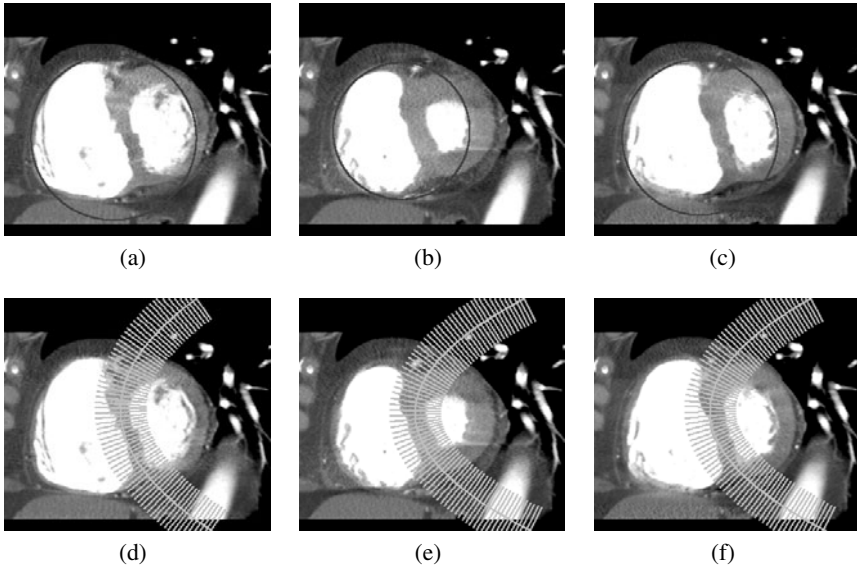


Fig. 5 Localization of interventricular septum: (a), (b), (c) - the smallest circular contour containing both ventricles, (d), (e), (f) - parabola approximately localizing the interventricular septum.

inside the contour. The weight w was set to 20, which was chosen experimentally. The results of that phase are shown in Fig. 5.

The circular contour found in the first phase was used to choose the parabola representing interventricular septum. For that purpose the set of parabolas was considered. The parabolas in that set were defined by 3 points. Two of them were lying on the line containing the upper and lower border of the image while the third was located on the horizontal line going through the center of the image. To allow calculations only a finite number of points on that lines was chosen and consequently a finite number of parabolas was considered. To find the optimal one each parabola was examined to check how many pixels representing blood inside the circular contours from the first phase is located under the considered curve. Additionally, since there can be many parabolas with the same number of such pixels, it was checked which one is equally distant from the pixels representing blood in the direction perpendicular to the given parabola. This process and the results is also presented in Fig. 5. It is worth mentioning that this procedure can be considered as a separate *active contour* method where searching for the optimal curve analyzes all the elements of contour set.

Similarly to results presented in the previous section, the proposed approach was also evaluated using 1200 images. This time, however, it was checked how many pixels, marked by an expert as pixels lying either in left or in right ventricle, lie

Table 2 Distribution of incorrectly classified pixels among 1200 analyzed images during interventricular septum localization.

	Number of pixels				
	[0, 100)	[100, 200)	[200, 300)	[300, 400)	[400, ∞)
Number of images	1151	42	3	0	4

on wrong side of the found parabola. The average number of such pixels was 11 from 89600 pixels in the image. Distribution of the incorrectly classified pixels is presented in Table 2.

5 Segmentation

The knowledge extracted in two previous sections was used to segment heart images using *potential active contours* described in [5]. In this work the sample results of left ventricle segmentation are presented. In the conducted experiments potential contour was defined using one object source and two background sources. Energy function was similar to the function described by equation (1). In this case, however, E^{out} component considered only pixels lying on the right side of the parabola representing interventricular septum and weight w was set to 10. The sample results of that approach are presented in Fig. 6.

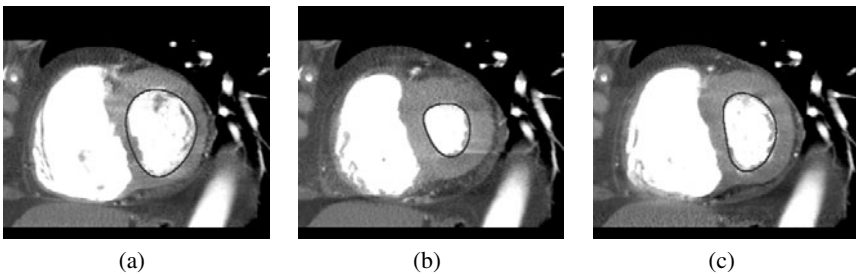


Fig. 6 Segmentation results of left ventricle using *potential active contours*.

6 Summary

In this paper a methodology of knowledge extraction, indispensable for heart ventricle segmentation in computed tomography (CT images, was described. The proposed approach allowed to identify blood inside ventricles as well as an approximate localization of interventricular system. Results presented in sections 3 and 4 reveal that the method used gives very good effects, which is additionally confirmed by the segmentation results that were shown in section 5. The advantage of the presented

work is fact that the extracted knowledge can be of use in any other system of image understanding ([4, 7]) analyzing heart images.

Acknowledgement. Authors would like to express their gratitude to Mr Cyprian Wolski, MD, from the Department of Radiology and Diagnostic Imaging of Barlicki University Hospital in Lodz for making heart images available and sharing his knowledge.

References

- [1] Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *International Journal of Computer Vision* 22(1), 61–79 (2000)
- [2] Grzeszczuk, R., Levin, D.: Brownian strings: Segmenting images with stochastically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(10), 1100–1113 (1997)
- [3] Kass, M., Witkin, W., Terzopoulos, S.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–333 (1988)
- [4] Tadeusiewicz, R., Ogiela, M.R.: *Medical Image Understanding Technology*. STUD-FUZZ, vol. 156. Springer, Heidelberg (2004)
- [5] Tomczyk, A.: Image segmentation using adaptive potential active contours. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.) *Proceedings of 5th International Conference on Computer Recognition Systems (CORES 2007)*. *Advances in Soft Computing*, pp. 148–155. Springer, Heidelberg (October 2007)
- [6] Tomczyk, A., Wolski, C., Szczepaniak, P.S., Rotkiewicz, A.: Analysis of changes in heart ventricle shape using contextual potential active contours. In: Kurzynski, M., Wozniak, M. (eds.) *Proceedings of 6th International Conference on Computer Recognition Systems (CORES 2009)*. *AISC*, pp. 397–405. Springer, Heidelberg (2009)
- [7] Tomczyk, A., Pryczek, M., Walczak, S., Jojczyk, K., Szczepaniak, P.S.: Spatch based active partitions with linguistically formulated energy. *Journal of Applied Computer Science* 18(1), 87–115 (2010)

Estimation of Wandering ECG Signal Baseline Using Detection of Characteristic Points in PQRST Complex for Avoiding Distortions in Complex Morphology

Joanna Śledzik and Mirosława Stelengowska

Abstract. This paper presents an approach for wandering baseline correction in electrocardiographic signal for diagnostic purposes. It presents problems of distortions of ECG signal caused by environment and physiological reasons, particularly - the ECG baseline wandering problem. The main goal of this paper is to present method for dealing with ECG baseline wander - estimation (by B-spline interpolation) of baseline, developed by authors on the basis of cubic spline interpolation method introduced by Fabio Badilini [3] and extended by using additional detected PQRST characteristic points in order to get better fitness. The experiments with comparison to F. Badilini method are made on artificial signals with various heart rate and ST segment morphology parameters with respiration effect (the respiration mechanism is the main cause of ECG baseline wander).

1 Introduction

PQRST complex morphology – the shape and amplitude of particular waves – is the base for ECG diagnosis of various heart diseases. Figure 1 presents the schema of PQRST complex with marked waves. The biphasic shape of the first wave (P-wave) stands for atrial enlargement which can lead to serious distortions in heart contraction cycle resulting in fainting (the proportion between phases indicates which atria is enlarged). The amplitude of central R-wave, direction and shape of T-wave are the base for detection of ventricular enlargement [5]. The shift of ST segment is a mark of symptoms of myocardial ischemia - very

Joanna Śledzik

Institute of Medical Technology and Equipment, 41-800 Zabrze, ul. Roosevelta 118
e-mail: joanna.sledzik@itam.zabrze.pl

Mirosława Stelengowska

Institute of Medical Technology and Equipment, 41-800 Zabrze, ul. Roosevelta 118
e-mail: miroslaw.stelengowska@itam.zabrze.pl

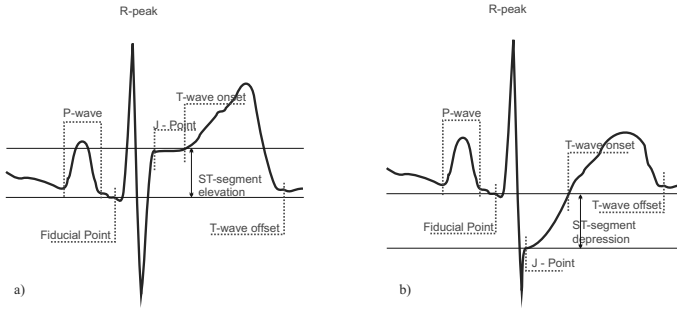


Fig. 1 Sample PQRST complexes in two variations - with elevation (a) and depression (b) of ST segment with marked waves and points significant for diagnostic purposes

dangerous state of imbalance between heart muscle oxygen demands and coronary arteries supply possibilities due to arteriosclerotic changes. That disease is at present one of the strongest factors in adult mortality (it leads to acute myocardial infarction and can result in sudden cardiac death). The shift of ST segment allows to assess the size and location (the afflicted region of heart muscle) of myocardial ischemia [8]. Therefore, it is very critical task to remove distortions in ECG signal without deforming the waves morphology. Unfortunately, distortions during acquiring ECG signal from body surface are very significant. They mainly originate from following physiological and environmental reasons: energy network, acting impulses in chest muscles and respiration mechanism (the subsection 2.1 *ECG filtering* explains in detail these distortions). Development of the ECG filtering mechanisms is one of the critical research tasks in ECG analysis. The main difficulty is that the more effective filter is, the stronger changes it cause in the original shape (morphology) of PQRST complexes and can lead to false diagnostic conclusions or hide the symptoms of one of diseases mentioned above. There have been developed many regulations for filter characteristics for energy network filter and highpass filter [7]. This paper deals with the distortion whose filtration strongly modifies PQRST morphology, especially ST segment responsible for showing myocardial ischemia symptoms – the ECG baseline wandering. Figure 2 present baseline wandering effect and Figure 3 - the desired effect of wandering baseline elimination. The significance of distortions caused by classical filter forces the researchers to search the alternative approach and the one of the best results was estimation of original wandering baseline by cubic spline estimation introduced by Fabio Badilini in his work *Cubic spline baseline estimation in ambulatory ECG recording for the measurement of ST segment displacements* [3]. Fabio Badilini used the fiducial points of every PQRST (see Figure 1) to create, by cubic spline interpolation, parallel signal containing wandering baseline which was finally subtracted from the original signal. Other

researches have proved that replacement of baseline filter with baseline interpolation brings positive results, especially in beat-to-beat ST segment analysis[2] in Holter recordings and microvolt T-wave alternans analysis[4]. Experiments have shown that one point for QRS can be insufficient to exact fitting and authors decided to extend this method by adding other points from QRS that lie on isoline and can be detected on the basis of first derivative – T-wave offset and isoline between T-wave offset and fiducial point of next QRS (but not in the region of P-wave). The subsection 2.2 *Alternate dealing with baseline wander* presents the basics of the detection and interpolation of authors approach. The results, presented in section 3, are quite promising and show that spline interpolation using mentioned three points from two following PQRST complexes leads to better fitness. The experiments were performed on the set of artificial signals generated by multiparameter simulator, being combination of extremely low (40 beats per minute - BPM), medium (80 BPM) and high (180 BPM) heart rate and ST segment - small (0.2 mV) and high (0.8 mV) - elevation/depression. The accuracy of fitness is assessed by root mean square error (RMSE) in signals first derivative (which reflects the shape of signal and is not influenced by the absolute level of the isoline). The testing set consisted of the signals with added respiration effect and the reference set consisted of the same signals, but without respiration effect. Section 4 presents conclusions and plans for further development.

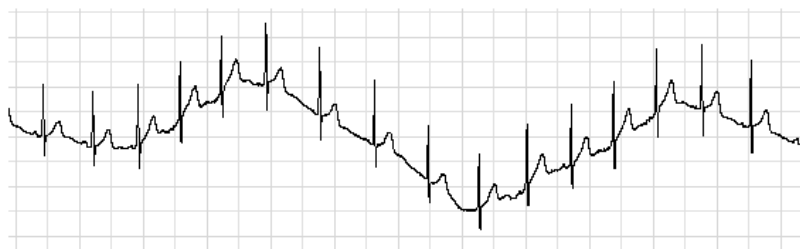


Fig. 2 Effect of baseline wander due to respiration



Fig. 3 Filtered signal ready to analysis

2 ECG Signal Processing and Methods of Baseline Wandering Handling

2.1 ECG Filtering

ECG signal that is gathered from body surface contains following distortions:

- energy network, the sinusoid with fixed frequency (50Hz in Poland),
- muscle, coming from chest muscles,
- low-frequency distortions (including baseline wander due to respiration mechanism).

Energy network distortions have fixed, known frequency and are removed by notch filters. There exists significant difficulty in construction of such filter - Polish Standards for power industry allows the deviation of energy frequency between 49–51 Hz. Muscle distortions are attenuated by single or cascade of moving average filters. The third, low-frequency distortions are reduced using highpass filter. This filter distorts ST segment morphology and there exist official, published as a standard, recommendations for characteristics of such filter: -3dB at 0.67 Hz for vital monitoring purposes and -0.9 dB at 0.67 Hz - for diagnostic purposes [7].

Filtering of baseline wander becomes a serious problem for signal original morphology, especially for ST segment. Unfortunately - the more effective filter is, the stronger ST-segment distortions it causes. Figure 4 presents the deformations caused by applying baseline wander filter. Original signal is an artificial signal with following parameters: normal PQRST morphology, heart rate 60 BPM and ST elevation 0.2 mV. The second signal is the same signal, however with added respiration artifact and filtered with high-pass filter.

2.2 Alternate Dealing with Baseline Wander

The key to avoid using the baseline wandering filter is to estimate the baseline and correct the amplitude of necessary ECG samples to perform ST segment level measurements. Fabio Badilini in his work *Cubic spline baseline estimation in ambulatory ECG recording for the measurement of ST segment displacements* [3] presents approach for estimation through interpolating consecutive isoelectric points with third order lines. First, fiducial points correlated with every QRS are detected by local minimum detection in 100 ms length window before R-peak and then, the consecutive four of them are used in cubic spline interpolation to acquire isoline third order function. Authors made an attempt to extend the set of points in each PQRST complex to get better fitness. There were identified points that, on the basis of physiology, lie on the level of isoline: mentioned fiducial point, T-wave offset and point between T-wave offset and fiducial point of following complex (due to interpolation mechanism there was chosen a point that lies approximately in the middle of them, but not in range of P-wave). This approach required algorithm for detection

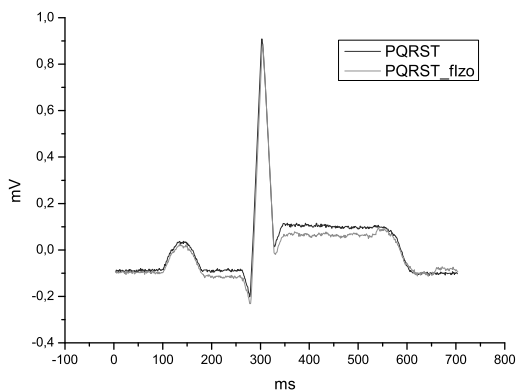


Fig. 4 The comparison of artificial PQRST complexes with 0.2 mV ST elevation: first, recorded without baselin wander artifact (PQRST) and second, recorded with baseline wander artifact and filtered (PQRST_fIzo)

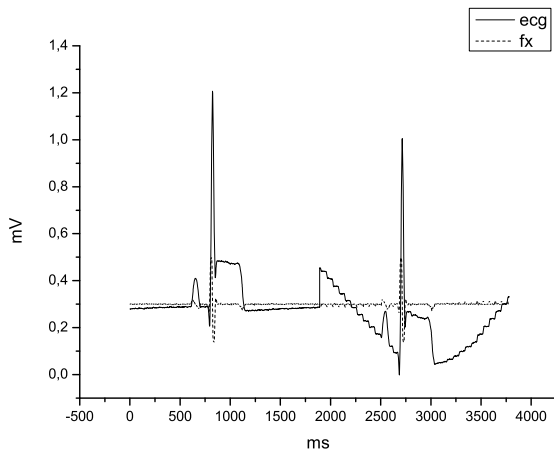


Fig. 5 Sample PQRST complexes from signal without and with baseline wander with its first derivative, normalized for presentation

of P-wave and T-wave offset. In this paper authors decided to choose first derivative for the base for detecion of R-peak and the waves. In case of P-wave first derivative has got characteristic, two-phased shape and at T-wave offset – one-phased, as presented at Figure 5. The detection algorithm implements detection of such patterns in surroundings of previously detected R-peak.

As it can be observed, the shape of the first derivative in P-wave and T-wave regions does not significantly change with adding respiration effect what makes this method suitable for this application. Finally there is performed interpolation in steps where one step covers following two PQRST complexes with set of seven points: point on isoline before PQRST complex No. 1, its fiducial point and T-wave offset, point on isoline between PQRST complex No. 1 and 2, fiducial point and T-wave offset of PQRST No. 2 and point of isoline after it. In online mode, which requires in-time detection the last value is created on a basis of RR-distance of current and previous PQRST. The next section presents the results of the algorithm on set of artificial ECG signals with varying heart rate and ST elevation/depression level.

3 Results

The input signals were the artificial ECG signals from Fluke Biomedical MPS 450 Multiparameter Simulator with varying heart rate and ST elevation/depression parameters with added respiration artifact. authors chose two extreme heart rate value - 40 BPM (bradycardia), 180 BPM (tachycardia) and middle 80 BPM (normocardia). Then signal with each heart reate was recorded with four different ST segment levels: 0.2 mV/0.8 mV elevation and 0.2 mV/0.8mV depression. Final set consisted with 12 signals with added respiration effect (test portion) and without respiration effect (reference portion). Signal was gathered by certified ECG module filtered with built-in analog highpass filter 0.05 Hz, conforming the IEC standard [7] and filtered with software 49/51 Hz notch filter and cascade moving average filter with small radius: 5 samples. The sample rate was 1000 Hz, the length of signal portion was 30 ms. There was applied R-peak detection algorithm based on the first derivative (first-derivative based QRS detection methods are the most popular and time-effective ones [1]). Authors applied Holsinger algorithm [6] which assumes searching 12 ms signal portion which exceeds fixed threshold. However, the pure Holsinger algorithm has a high incidence of false positives, it is often extended with refraction period which helps to recognize false positive results which, in the basis of human heart rate limits, are surely beetwen R-peaks [9]. In this paper authors fixed refraction period to 150 ms.

Then alternatively was applied:

- cubic spline interpolation by Fabio Badillini,
- algorithm created by authors.

Each algorithm resulted with parallel signal containing wandering baseline which was finally substracted from signal with respiration effect. authors decided to assess and compare results using root mean square error (RMSE) between first derivatives of test and reference signals. Using RMSE beetwen samples value leded to assesment errors, because reference signals baseline does not have 0 mV amplitude and such variable significantly grants algorithm which creates baseline on absolute level

Table 1 The comparison of measurement errors between Fabio Badilini cubic spline interpolation and interpolation developed by the authors

HR	ST	Badilini RMSE	authors RMSE
40 BPM	1. ST +0.2mV	0.0027	0.0009
	2. ST +0.8mV	0.0028	0.0011
	3. ST -0.2mV	0.0012	0.0009
	4. ST -0.8mV	0.0014	0.0009
80 BPM	5. ST +0.2mV	0.0014	0.0012
	6. ST +0.8mV	0.0015	0.0013
	7. ST -0.2mV	0.0161	0.0160
	8. ST -0.8mV	0.0011	0.0009
180 BPM	9. ST +0.2mV	0.0027	0.047
	10. ST +0.8mV	0.0088	0.0121
	11. ST -0.2mV	0.0011	0.0040
	12. ST -0.8mV	0.0078	0.0171

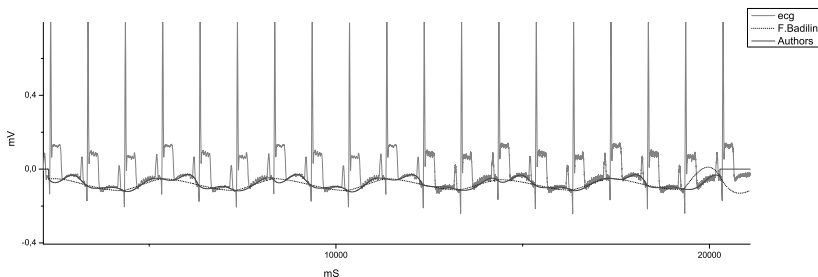


Fig. 6 Sample results of baseline estimation on signal with heart rate 80 BPM and +0.2mV ST elevation - recorded ECG signal (ecg) with superimposed baseline interpolated using F.Badilini and the authors method

closer to original - in practice it less important than proper shape of signal. Additionally, the parameters of PQRST waves in diagnostic purposes are always measured and analysed using relative values so the baseline absolute level isn't strongly significant. The results are presented in Table 1. Figure 6 graphically presents sample results on one of the signal sets.

As it can be observed - authors algorithm gives better results in all the cases - except the last four signals with extremely high heart rate. Both algorithms are less effective in this case, but authors algorithm gives worse results than F.Badilini's one. The following figures show in details (on only four PQRST complex) baseline estimation effects on normal heart rate (Figure 7) and tachycardiac rate (Figure 8).

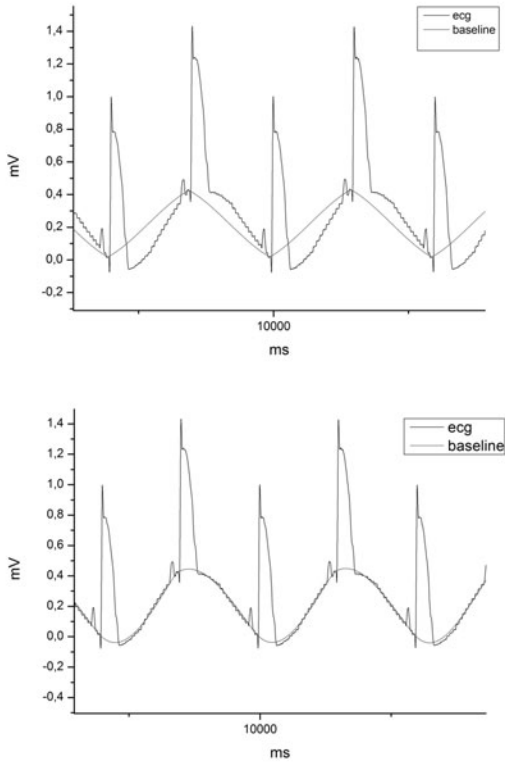


Fig. 7 Comparison of baseline estimations made by: F.Badilini method (top chart) and the authors (bottom chart) on signal with heart rate 40 BPM and 0.8 mV ST segment elevation

In normal heart rate (Figure 7) it can be observed that baseline estimated only with fiducial points is very simplified (on signal with 40 BMP heart rate it closes to a triangular shape) while the baseline estimated by authors method is well shaped. In extremely high heart rate (Figure 8) baseline estimated by F.Badilini method on the top chart has well shape. On the bottom chart (baseline estimated by the authors method) it can be observed negative effect of overfitting that is responsible for high RMSE error presented in Table 1.

However, the fact that such negative effect exists only in high heart rates allows to construct the method joining the F.Badilini and authors approaches. Authors approach includes all issues necessary for F.Badilini method and in online mode it is possible to simply switch to the latter when heart rate significantly increases.

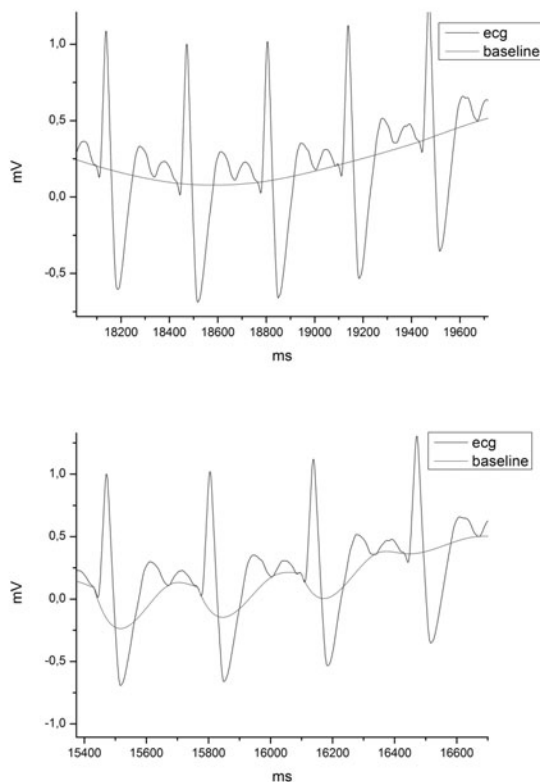


Fig. 8 Comparison of baseline estimations made by: F.Badilini method (top chart) and the authors (bottom chart) on signal with heart rate 180 BPM and 0.8 mV ST segment depression

4 Summary

The baseline interpolation using three points per PQRST gives promising results. It is especially valuable in beat-to-beat PQRST analysis in Holter recordings [2] which - in opposite to stress exercise test - often consist of slight changes in low frequency components - especially in ST segment. Furthermore, recent researches show that baseline estimation in opposite to baseline filters is very useful in T-wave alternans analysis [4] which allows to assess long-term risk of heart disease. The preservation of morphology is very important and developed baseline interpolation algorithm will be very useful. The plans for further development contain improvement of assessment in case of high heart rate and improvement of PQRST waves detection on the basis of rare and pathological cases of wave shapes.

Acknowledgements. This work was partially financed by the Ministry of Science and Higher Education resources under Research Project NR 13 0052 10.

References

- [1] Arzeno, N., Zhi-De, D., Chi-Sang, P.: Analysis of First-Derivative Based QRS Detection Algorithms. *IEEE Trans. on Biomedical Engineering* 55, 478–484 (2008)
- [2] Badilini, F., Merri, M., et al.: Cubic spline baseline estimation in ambulatory ECG recordings for the measurement of ST segment displacements. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 13, pp. 584–585 (1991)
- [3] Badilini, F., Moss, A., Titlebaum, E.: Beat-to-Beat Quantification and Analysis of ST Displacement from Holter ECGs: A New Approach to Ischemia Detection. In: *Proc. of Computers in Cardiology*, pp. 179–182 (1992)
- [4] Burattini, L., Zareba, W., Burattini, R.: Automatic detection of microvolt T-wave alternans in Holter recordings: Effect of baseline wandering. *Biomedical Signal Processing and Control* 1, 162–168 (2006)
- [5] Dubin, D.: Rapid interpretation of EKG's. In: *PZWL, Warsaw*, pp. 243–259 (2000)
- [6] Holsinger, W., Kempner, K., et al.: A QRS preprocessor based on digital differentiation. *IEEE Trans. on Biomed. Eng.* 18, 212–217 (1971)
- [7] IEC 60601-2-51:2003. Medical electrical equipment. Part 2-51: Particular requirements for safety, including essential performance, of recording and analysing single channel and multichannel electrocardiographs
- [8] Maclachlan, M., Nielsen, B.: Computing the size and location of myocardial ischemia using measurements of ST-segment shift. *IEEE Trans. on Biomed. Eng.* 53, 1024–1031 (2006)
- [9] Vidal, C., Charnay, P., Arce, P.: Enhancement of a QRS detection algorithm basen on the first derivative, using techniques of a QRS detector algorithm based on non-linear transformations. In: *Proc. of 4th European Conference of the International Federation for Med. and Biol. Eng. IFMBE*, vol. 22, pp. 393–396 (2008)

Continuous Wavelet Transform as an Effective Tools for Detecting Motion Artifacts in Electrogastrographical Signals

Barbara T. Mika, Ewaryst J. Tkacz, and Paweł S. Kostka

Abstract. The cutaneous recording of gastric myoelectrical activity of the stomach known as electrogastrigraphy (EGG) seems to be the promising tool for the non-invasive assessment of gastric motility. Unfortunately the EGG recording is usually severely contaminated both by motion artefacts and endogenous biological noise source. In order to use EGG signals as reliable diagnostic tool it is necessity to look for the effective artefacts removal methods. In this paper Continuous Wavelet Transform (CWT) was applied for detection motion artefacts from the EGG data. The set of own mother wavelets extracted directly from EGG signal was created and applied for detecting motion artefacts from one channel EGG recording. The results was compared with the effects obtained by using standard mother wavelets. The proposed method based on CWT with own mother wavelet presents very good performance for detecting motion artefacts from the EGG data.

1 Introduction

Electrogastrogram (EGG) reflects the myoelectrical activity of stomach picked up by the electrodes placed cutaneously on the abdominal skin over the stomach. As a non-invasive test, easy to perform and relatively inexpensive EGG has been an

Barbara T. Mika

Silesian University of Technology, Institute of Electronics,
Division of Microelectronics and Biotechnology, Gliwice, Poland
e-mail: b_mika@wp.pl

Ewaryst J. Tkacz

Silesian University of Technology, Institute of Electronics,
Division of Microelectronics and Biotechnology, Gliwice,
Academy of Business, IT Department Manager in Dąbrowa Górnicza, Poland
e-mail: etkacz@polsl.pl

Paweł S. Kostka

Silesian University of Technology, Institute of Electronics,
Division of Microelectronics and Biotechnology, Gliwice, Poland
e-mail: pkostka@polsl.pl

attractive tool for physiological and pathophysiological studies of the stomach [2]. For the first time such measurement was carried out in 1922 by Walter Alvarez [1], unfortunately the lack of both high technology of registration signals and advanced method of signal processing was the main causes that delayed medical application of EGG in gastrointestinal diseases treatment. Clinical benefits from information encoded in the EGG recording highly depend on the quality of extracted signal. The main component of gastric myoelectrical activity, called gastric slow waves, has a frequency about 3 cycle per minute (0.05 Hz). Similar to the other biological signals, EGG apart from mentioned normal physiological rhythm 2-4 cpm (0.033-0.066 Hz) called normogastric rhythm, includes some additional pathological rhythms covering frequencies from 0.5 cpm up to 9 cpm (0.008-0.15 Hz). Due to the leading rhythm in the EGG signal it is possible to distinguish: bradygastric rhythm [0.5, 2] cpm ([0.01, 0.033] Hz) normogastric rhythm [2,4] cpm ([0.033, 0.066] Hz) and finally tachygastric rhythm (4, 9] cpm ([0.066, 0.15] Hz) [5]. Compared with the other electrophysiological measurements, the EGG has a low signal-to-noise ratio and is usually contaminated by cardiac, respiratory signal and possible myoelectrical activity from other organs situated near the stomach such as duodenal or small intestine. As EGG signal require at least half an hour recording before stimulation and the same time after stimulation so it is almost impossible to avoid motion artefacts which are a specially strong as far as children are concerned. Artefacts damage the recorded data and make interpretation of EGG very difficult or even impossible. Advanced signal processing methods are needed for studying the gastric electrical behavior. In order to use EGG as a diagnostic tool the artefacts have to be detected and automatically eliminated before analysis.

1.1 Materials and Methods

The multiresolution signal decomposition methods analyze signals at different scales or resolution and thus ideally suit for studying the non-stationary signal such as EGG [3],[4],[7]. Multiresolution representation is computed by decomposing original signal using wavelet orthogonal basis and could be treated as decomposition using set of independent frequency channels [6]. The wavelet transform has varying window size that provide optimal time-resolution in all the frequency ranges. As the Continuous Wavelet Transform (CWT) exhibits the property of "zooming" in the sharp temporal variation in a signal the object of this study is to examine the applicability of CWT for detecting motion artefacts in the EGG signals.

The CWT of one-dimensional signal $f(t) \in L^2(R)$ with respect to a mother wavelet $\psi(t)$ is defined as :

$$W_{\psi}\{f(t)\}(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt \quad (1)$$

where a, b ($a, b \in R \wedge a \neq 0$) are the scale and translation parameters, respectively.

The idea of using wavelets for detection motion artefacts in the one-channel EGG recording, based on cutting the part of original signal, which after visual inspection was decided to be an artefact, and create an appropriate wavelet ψ which the best "matches" the waveshape of extracting artefact. After visual inspection of various motion artefacts shapes, raw one-dimensional EGG signal recording from four different persons were taken under consideration. From the parts of different original signals, which were qualified by the physician, as the motion artefacts, were created four (similar in the shape with distinguished artefacts) mother wavelets, named adequately: egg1, egg2, egg3, egg4.

Figure 1 displays two hours recording of one-channel EGG (Mal 61), 30 minutes in fasting state and 90 minutes after stimulation with 300 ml of 237 kcal energy value yogurt. Ten seconds lasting motion artefact which appeared between 89 minutes 55 seconds and 90 minutes 05 seconds pointed out with an arrow in the graph (fig.1.) was the source for constructing the mother wavelet egg3. It is shown in the fig.2 that constructed wavelet egg3 is the same in the shape as the indicated artefact.

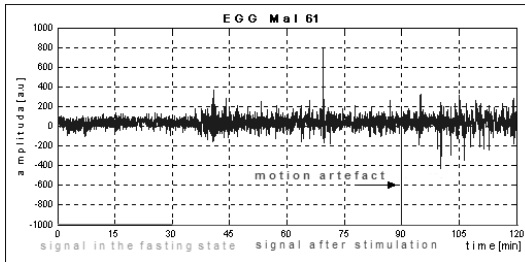


Fig. 1 Two hours recording of one channel EGG signal Mal61 with motion artefacts.

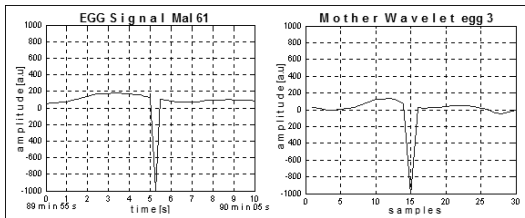


Fig. 2 The zoom of 10 seconds motion artefact in signal Mal61 pointed out by an arrow in fig.1 next to created mother wavelet egg3.

For detecting motion artefacts in the EGG signal four mother wavelets (egg1, egg2, egg3, egg4) has been constructed (fig.3). Each of them were created on the basis of part of the one channel EGG signal recorded from different persons (egg1-signal Mal57, egg2-signal Mal60, egg3-signal Mal61 and egg4-signal Mal63). Two mother wavelets (egg1, egg2) after interpolation consist of 20 samples and the others include 30 samples.

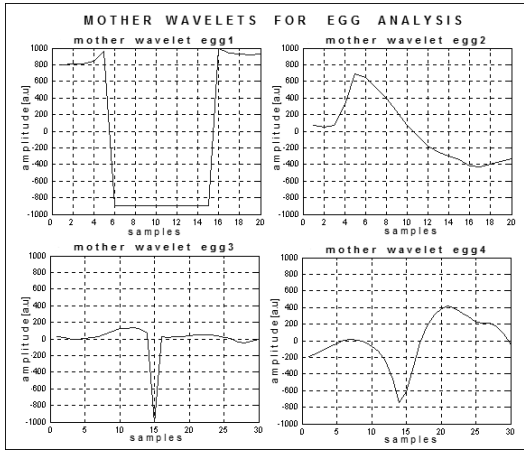


Fig. 3 Set of constructed mother wavelets : egg1, egg2, egg3, egg4 for EGG analysis.

The CWT stated mathematically by the formula (1) was applied for one channel EGG signal. In this case function f is the raw EGG signal which should be cleaned from motion artefacts, and the function $\psi = \text{eggi} \wedge i \in \{1, 2, 3, 4\}$ for $a = b = 1$ is one of created or standard mother wavelets translated by b and scaled by a . Both parameters in this study have been set to one.

Using CWT as a tool for inspecting the EGG signal, totally control correlation process was obtained, it means that CWT coefficients reach the maximum value for this part of original signal that, is the most similar to the wavelet used for analysis. CWT coefficients are the key for localization motion artefacts in the examining signal.

The EGG analysis was conducted by the aid of "wavemngr" library of *Wavelet Toolbox in Matlab*. The results obtained by applying CWT with constructed four mother wavelets were compared with the results obtained for some standard mother wavelets such as: Morlet, Daubechies-3, Discrete Meyer and Gauss-2. At the first step for each one-channel EGG signal examination, the length (in samples) of motion artefact in the EGG signal was estimated manually by the aid of Signal Processing Toolbox of Matlab, as reference data. Afterwards by applying CWT both with created and standard mother wavelets for 8 scales, each of studding signal were represented by $8 \times N$ matrix of wavelet coefficient (N - number of samples) for each used mother wavelet. Each row of wavelet coefficients matrix consist of N wavelet coefficients for each scale. The maximum value of CWT coefficient point out the localization (first sample) for detected motion artefact.

1.2 Results

The results from the table 1 was shown on the time representation of EGG Mal64 (fig.4) where the markers indicate the beginning and the end of motion artefacts.

Table 1 The below table presents the samples number for automatically detected motion artefacts by the aid of CWT. The first sample number for two detected motion artefact was underline.

	Created Mother Wavelets				Standard Mother Wavelets			
	Sample number of detecting artefact							
scale nb.	egg1	egg2	egg3	egg4	morl	db3	dmey	gauss2
1.	3901	11890	3868	<u>3870</u>	2	1	1	2
2.	<u>11900</u>	3869	3863	3871	3	2	2	2
3.	<u>3885</u>	3869	3864	3874	2	1	2	12098
4.	3887	3870	3864	3875	2	2	2	11884
5.	3892	<u>3867</u>	3864	3887	2	3876	3878	126
6.	3895	11876	3861	11893	3876	3879	3878	127
7.	3898	11876	<u>3860</u>	<u>11891</u>	3878	<u>3873</u>	3878	<u>11900</u>
8.	3901	<u>11873</u>	3861	11895	<u>3873</u>	3873	<u>3873</u>	<u>3873</u>

EGG signal Mal 64

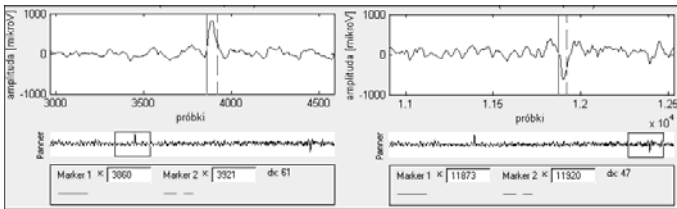


Fig. 4 Signal Mal 64 with two motion artefacts detected by CWT. The markers were set according the samples number underline in the table1 (for the last sample number the length of mother wavelet used for analysis was added). The first artefact last 8 seconds the other one 6 seconds.

1.3 Conclusions

In fig.5, mother wavelets egg1 and egg2 clearly pointed out two motion artefacts for each scale (the first artefact started at 3860 sample and the second one at 11873), while both Morlet and Daubechies-3 wavelet detection is weaker especially as far as the second artefact is concerned (started at 11873 sample). Similarly in fig. 6 the detection with created mother wavelets egg3 and egg4 is also better. The further studies which were conducted for different EGG data, using CWT with mother wavelet, created as we depict in this paper, confirm that all constructed wavelets are more successful in recognizing and localization motion artefacts then the standard ones. The proposed procedure is also easily applicable for multichannel EGG data, where it gives the opportunity for insight in the EGG propagation. Presented

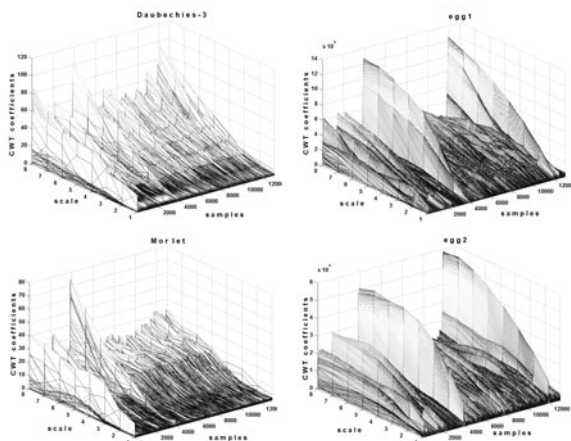
EGG Mal 64

Fig. 5 Motion artefacts localization in EGG signal Mal64. Results obtained after applying CWT, on the left side with the standard mother wavelets: Daubechies-3 & Morlet, on the right side with created mother wavelets: egg1, egg2.

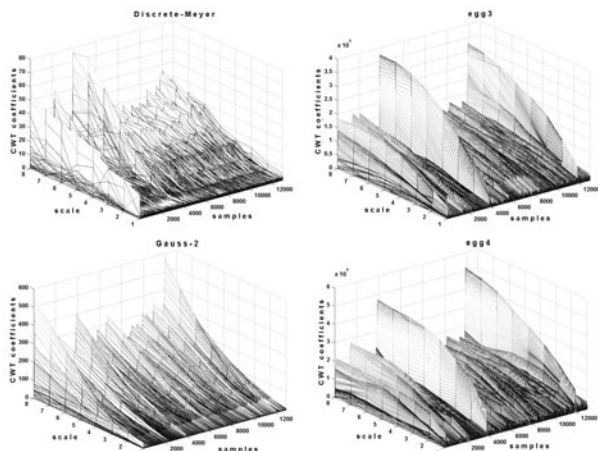
EGG Mal 64

Fig. 6 Motion artefacts localization in EGG signal Mal64. Results obtained after applying CWT, on the left side with the standard mother wavelets: Discrete Meyer & Gauss-2, on the right side with created mother wavelets: egg3, egg4.

method of detection motion artefacts from EGG recording seems to be very promising tool for effective application, but needs further investigation in order to increase its preciseness.

Acknowledgements. This work has been supported by the Ministry of Science and Higher Education in the years 2010-2012 as development project O R00 0132 12.

References

- [1] Alvarez, W.C.: The Electrogastragram and What It Shows. *Journal of the Americal Medical Association* 78, 1116–1119 (1992)
- [2] Koch, K.L., Stern, R.M.: *Handbook of Electrogastrography*. Oxford University Press, Oxford (2004)
- [3] Liang, H., Lin, Z.: Stimulus Artifact Cancellation in The Serosal Recordings of Gastric Myoelectrical Activity Using Wavelet Transform. *IEEE Transactions on Biomedical Engineering* 49(7) (July 2002)
- [4] Liang, H., Lin, Z.: Multiresolution Signal Decomposition and Its Applications to Electrogastric Signals. *Recent Research Developments in Biomedical Engineering* 1, 15–31 (2002)
- [5] Parkman, H.P., Hasler, W.L., Barnett, J.L., Eaker, E.Y.: *Electrogastrography: a Document Prepared by The Gastric Section of The American Motility Society Clinical GI Motility Testing Task Force*. *Neurogastroenterol Motility* 15, 89–102 (2003)
- [6] Mallat, S.G.: A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis Machine Intelligence* 11(7) (1989)
- [7] de Sobral Cintra, R.J., Tchervensky, I.V., Dimitrov, V.S., Mintchev, M.R.: Optimal Wavelets for Electrogastrography. In: *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEMBS 2004* (2004)

Asymmetry of Digital Images Describing Melanocytic Skin Lesions

Pawel Cudek, Jerzy W. Grzymala-Busse, and Zdzislaw S. Hippe

Abstract. This paper presents a method for automatic identification of asymmetry in digital images containing melanocytic skin lesion. Our method is a part of the new system for classifying skin lesion using Stolz strategy, based on the primary ABCD rule.

1 Introduction

Available literature [1] shows that melanoma (Melanoma malignant, MM) is a serious threat to human life and health. Moreover, the number of occurrences of this type of skin cancer is increasing every year. Therefore early and appropriate classification of this type of lesion is very important. This situation inspired us to extend the **I**nternet **M**elanoma **D**iagnosing and **L**earning **S**ystem (**IMDLS**) [2, 3] with diagnostic module based on digital medical image analysis of actual lesions occurring on skin. Our new tool can be a supplement of diagnosing process and may facilitate suitable medical procedures, giving an indication for the necessity of the lesion's surgical removal.

Melanocytic skin lesion classification and related hazard degree assessment are generally done by dermatologists using certain standard procedures (often called melanocytic algorithms, or strategies). Generally all known strategies [4, 5, 6], involve a human or a suitable machine detecting certain characteristic features of the analysed lesion and indicating on that basis the necessity of surgery: *to excise* the lesion.

Pawel Cudek · Jerzy W. Grzymala-Busse · Zdzislaw S. Hippe
University of Information Technology and Management, Rzeszow, Poland
e-mail: PCudek,ZHippe@wsiz.rzeszow.pl

Jerzy W. Grzymala-Busse
University of Kansas, Lawrence KS 66045, USA
e-mail: Jerzy@ku.edu

In our research for lesion classification we use the most widely known Stolz algorithm, formally based on the primary **ABCD** rule [4], in which **A** (*Asymmetry*) shows a result of evaluation of lesion's asymmetry, **B** (*Border*) estimates the character of a rim of the lesion, **C** (*Color*) identifies number of colors (one or more, from six allowed) present in the lesion, and **D** (*Diversity of structures*) stands for the number of structures (one or more, from five allowed). Elements of the **ABCD** rule enumerate four main symptoms of an investigated lesion, and at the same time these elements are used to compute **TDS** (**T**otal **D**ermatoscopy **S**core) parameter [7] according to (1).

$$\text{TDS} = 1.3 * \text{Asymmetry} + 0.1 * \text{Border} + 0.5 * \sum \text{Color} + 0.5 * \sum \text{Diversity of Structure} \quad (1)$$

For example, lesion vector of ABCD values can be presented as follows:

- Asymmetry - **symmetric change**,
- Border - **0**,
- Color - **four** selected colors present in lesion,
- Diversity of Structure - **four** selected structures present in a lesion

For this case, the TDS is 4.0 and is calculated as follows:

$$\text{TDS} = 1.3 * 0 + 0.1 * 0 + 0.5 * (0 + 0 + 1 + 1 + 1 + 1) + 0.5 * (1 + 0 + 1 + 1 + 1) = 4.0$$

According to the TDS value, the analyzed lesion could be assigned to one of four accepted categories (classes) of melanocytic skin lesions, namely: Begin nevus, Blue nevus, Suspicious nevus or Malignant melanoma (Table 1).

Table 1 Classification of melanocytic skin lesions in dependence of **TDS**-value

TDS value	Lesion classification
TDS <4,76 and lack of color blue	Begin nevus
TDS <4,76 and color blue is present	Blue nevus
4,76 >= TDS <5,45	Suspicious nevus
TDS >= 5,45	Malignant melanoma

Acquisition of the ABCD parameters is not difficult from the standpoint of a medical specialist. However, the automation of this process is a great challenge. In this article we will focus on automated image analysis that allows the determination of the asymmetry occurrence (symmetric lesion, lesion with one-axial asymmetry and lesion with two-axial asymmetry). The examined areas were extracted from medical images in way described in our previous paper [8].

1.0.1 A Collection of Image Data

Accuracy assessment of asymmetry in melanocytic lesion by using the created tool has been tested on a set of images of selected types of lesion. The necessary skin lesion images were obtained among others from [9], making up a set of 53 images of lesion classified with the **ABCD** rule. To increase the testing set, each image was rotated by 90, 180 and 270 degrees. This procedure provides a set of 212 real images of lesion, completely different from the perspective of classification automation. Image division due to the asymmetry is presented in Table 2.

Table 2 The division of the set of images depending on the number of detected symmetry axis.

Logical value of A	Number of cases referred by a medical specialist
symmetric lesion	88
1-axial asymmetry	52
2-axial asymmetry	72

2 Asymmetry Type Assessment

In the **ABCD** rule asymmetry assessment involve information about number associated with symmetry lying in the lesion area. A logical value of the attribute A can be: **symmetric lesion** (there are two perpendicular axes of symmetry), **1-axial asymmetry** (there is only one axis of symmetry) and **2-axial asymmetry** (there is no axis of symmetry). The numerical values used in the calculation of **TDS** parameter for the above logical values are 0, 1 and 2 points, respectively. The elaborated algorithm evaluation of asymmetry (Figure 1.) is based on the analysis of the black and white image created as a result of segmentation, in which white dots belongs to lesion area and black dots represents an area of healthy tissue.

In the first step of developed algorithm we determine gravity of center (**GC**), which the supposed symmetry axes can pass through. Coordinate x of the **GC** is defined as the sum of the coordinates x belonging to the lesion area divided by the total number of points forming a lesion. Analogously value of y coordinate of **GC** is the sum of y coordinates of points forming a lesion. The next stage of the algorithm is to create an array containing the length of straights (radiuses) outgoing from the **GC** point with the angle in range from 0 to 359 degrees.

An example of the graphical representation of straights and their values is shown in Figure 2.

The main step of the developed algorithm is testing a line, that may be a candidate for the symmetry axis of lesion. For this purpose, for each of the 180 potential candidate for the symmetry axis we compute the value of **SFA α** (**S**core **F**or **A**xis) is computed as presented in teh following procedure:

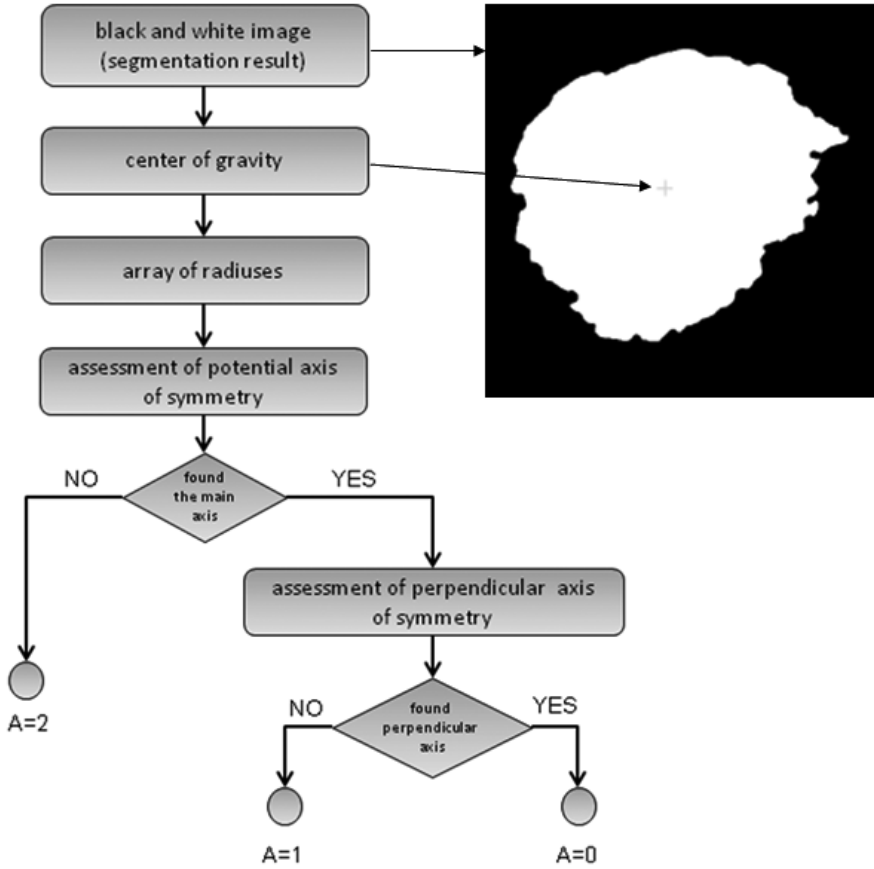


Fig. 1 Diagram of algorithm for the evaluation of asymmetry

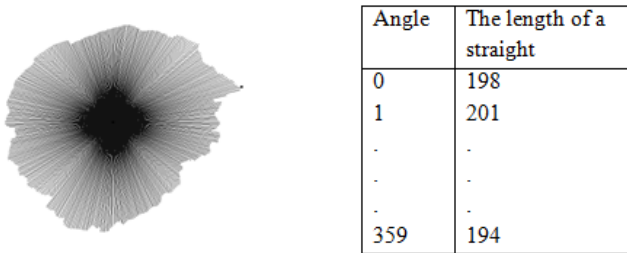


Fig. 2 Computed length of straights (radiuses).

Repeat for $\beta=1$ to $\beta=179$

$r1 = \text{get_radius}(\beta)$

$r2 = \text{get_radius}(-\beta)$

$\text{difference} = |r1 - r2|$

if($r1 < r2$)

$\text{difference_in_percentage} = \text{difference} / r1$

else

$\text{difference_in_percentage} = \text{difference} / r2$

if($\text{difference_in_percentage} \leq \text{radius_difference_threshold}$)

$SFA\alpha = SFA\alpha + 1$

where:

β - angle from 1 to 179 relative to the axis of symmetry

$SFA\alpha$ - score for the axis with an angle α

$r1, r2$ - radius

Therefore in this step the pairs of radiuses inclined to the tested axis at angles β and $-\beta$ are compared. For each pair of radiuses the difference in length, which is converted in the percentage of the length of the shorter radius is calculated. If the difference expressed as a percentage does not exceed the maximal value of 10% accepted tolerance, the rays are considered as similar and increase the number of votes "yes" ($SFA\alpha$) for the symmetry of the axis. Example of comparing each pair of rays shows Figure 3.

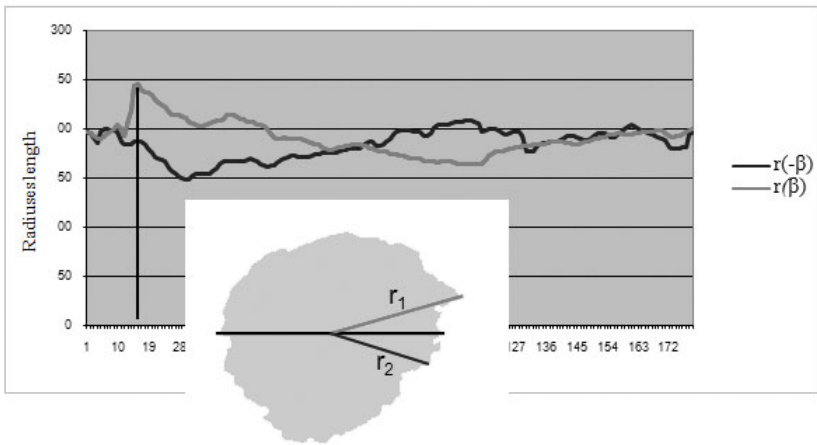


Fig. 3 Comparison of radiuses for the potential axis of symmetry with an angle α equal 0

The main axis of symmetry of the lesion is such for which the **SFA** is the largest and exceeds the threshold value agreed in researches (indicating an axis as the symmetry axis).

In the example shown in Figure 3 the main symmetry axis is straight with angle $\alpha=29^\circ$ for which 160 pairs of radiuses were rated as comparable. After detecting the main axis of symmetry, the **SFA** value of the perpendicular axis is checked. For this purpose the angle of the perpendicular axis is calculated ($\alpha + 90$) and the $SFA_{(\alpha+90)}$ value is searched in the table containing **SFA** for all angles. If the perpendicular axis can also be regarded as the axis of symmetry, the value of the parameter **A** in the ABCD rule is **0** (symmetric lesion). The failure of this test means that the parameter **A** is equal to **1** (1-axial asymmetry), and when it doesn't find the main axis of symmetry of the parameter **A** it is equal to **2** (2-axial asymmetry).

3 Results

In our research data set of images has been analyzed, using the developed algorithm (see Table 3). Results of classification gathered while reducing the number of radiuses to 180 and 90 has been also checked. This procedure reduces time needed to perform the calculations, however, significantly worse outcome classification. Therefore it seems appropriate to study the maximum number of radiuses.

Table 3 Percentage of correctly classified images depending on the number of radiuses.

Value of A	Number of radiuses		
	360	180	90
Symmetric lesion	94%	86%	82%
1-axial asymmetry	92%	79%	64%
2-axial asymmetry	91%	81%	83%

4 Conclusion

The developed algorithm is a next step in a way to create the computer diagnostic system supporting classification of melanocytic lesions based on digital images analysis.

Acknowledgement. This research has been supported by the grant No. N N516 482640 from the National Research Center.

References

- [1] Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., Smigal, C., Thun, M.J.: Cancer Statistics 2006 CA: A Cancer Journal for Clinicians 56(1), 106–130 (2006)

- [2] Grzymała-Busse, J.W., Hippe, Z.S., Knap, M., Paja, W.: Infoscience Technology: The Impact of Internet Accessible Melanoid Data on Health Issues. *Data Science Journal* 4, 77–81 (2005)
- [3] <http://www.melanoma.pl/>
- [4] Stolz, W., Braun-Falco, O., Bilek, P., Landthaler, M., Burgdorf, W., Cagnetta, A.: Atlas of Dermatoscopy. In: Czelej Edit. Office, Lublin, Poland, p. 210 (2006) (in Polish)
- [5] Argenziano, G., Fabbrocini, G., Carli, P., De Giorgi, V., Sammarco, E., Delfino, M.: Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch. Dermatol.* 134, 1563–1570 (1998)
- [6] Menzies, S.W.: Surface microscopy of pigmented skin tumors. *Australas J. Dermatol.* 38, 40–43 (1997)
- [7] Braun-Falco, O., Stolz, W., Bilek, P., Merkle, T., Landthaler, M.: Das dermatoskop. Eine Vereinfachung der Auflichtmikroskopie von pigmentierten Hautveränderungen. *Hautarzt* 40, 131 (1990)
- [8] Cudek, P., Grzymała-Busse, J.W., Hippe, Z.S.: Melanocytic Skin Lesion Image Classification. Part I: Recognition of Skin Lesion. In: Proceedings of 3th International Conference on Human System Interaction, pp. 251–257 (2010)
- [9] Triller, R., Aitken, G., Luc, T.: Atlas of Dermatoscopy of Pigmented Skin Tumors (2008), <http://www.plessfr/dermatoscopie>

Fuzzy Clustering and Adaptive Thresholding Based Segmentation Method for Breast Cancer Diagnosis

Paweł Filipczuk, Marek Kowal, and Andrzej Obuchowicz

Abstract. The paper provides a preview of some work in progress on the computer system to support breast cancer diagnosis. The approach is based on microscope images of the FNB (Fine Needle Biopsy) and assumes distinguishing malignant from benign cases. Research is focused on two different problems. The first is segmentation and extraction of morphometric parameters of nuclei present on cytological images. The second concentrates on breast cancer classification using selected features. Studies in both areas are conducted in parallel. This work is mainly devoted to the problem of image segmentation in order to obtain good quality features measurements. Correct segmentation is crucial for successful diagnosis. The paper describes hybrid segmentation algorithm based on fuzzy clustering and adaptive thresholding. The automatic system of malignancy classification was applied on a set of medical images with promising results.

1 Introduction

Breast cancer is the most common cancer among women. The prognosis in breast cancer is strongly dependent on the disease development before any treatment is applied so the chance of recovery is a function of time of the detection of cancer. Modern medicine does not provide one hundred percent reliable, if possible cheap and at the same time non-invasive diagnostic methods for the diagnosis of breast pathology. As a result, in practice the important function acting in breast cancer diagnosis is the so-called triple-test, which is based on the summary of results of three medical examinations with different degrees of sensitivity and it allows to achieving high confidence of diagnosis. The triple-test includes self examination (palpation), mammography or ultrasonography imaging and fine needle biopsy [19]. Fine needle biopsy is collecting nucleus material directly from tumor for microscopic

Paweł Filipczuk · Marek Kowal · Andrzej Obuchowicz

University of Zielona Góra, Institute of Control and Computation Engineering

e-mail: [P.Filipczuk, M.Kowal, A.Obuchowicz}@issi.uz.zgora.pl](mailto:{P.Filipczuk,M.Kowal,A.Obuchowicz}@issi.uz.zgora.pl)

verification. Next, the material (collected cells) is examined using microscope in order to confirm or exclude the presence of cancerous cells. The present approach requires a deep knowledge and experience of the cytologist responsible for diagnosis. In short, some pathologists can diagnose better than others. In order to make the decision independent of the arbitrary factor, morphometric analysis can be applied. Objective analysis of microscopic images of cells has been a goal of human pathology and cytology since the middle of the 19th century. Early work in this area consisted of simple manual measurements of cell and nuclear size. Along with the development of advanced vision systems and computer science, quantitative cytopathology has become a useful method for the detection of diseases, infections as well as many other disorders. In the literature one can find approaches to breast cancer classification [6, 8, 9, 10, 20]. Mentioned approaches are concentrated on classifying FNA (Fine Needle Aspiration) or FNB (Fine Needle Biopsy) slides as benign or malignant.

In this work, we present a fully automatic method that allows distinguish malignant cells from the benign cells. The classification of the tumor is based on morphometric examination of cell nuclei. In contrast to normal and benign nuclei, which are typically uniform in appearance, cancerous nuclei are characterized by irregular morphology that is reflected in several parameters. The best classification rate was achieved for such features as size, cell distribution on the image and color changes within the nuclei. It was decided not to use shape features because previous work showed that shape factors do not have good discriminative properties [11]. Features were extracted from segmented images obtained by hybrid segmentation method based on fuzzy clustering and adaptive thresholding described further in the paper.

The quality of segmentation and feature extraction was tested by using the set of classifying algorithms. The measure is based on classification accuracy obtained by leave-on-out cross-validation. In this work four different classification methods was used to rate the feature subsets: k-nearest neighbor, naive Bayes classifier, decision trees and classifiers ensemble [3, 12].

The paper is divided into four sections. Section 1 gives an overview of breast cancer diagnosis techniques. Section 2 describes the process of acquisition of images used to breast cancer diagnosis. Section 3 deals with segmentation algorithm used to separate cells and extract features. Section 4 shows the experimental results obtained using the proposed approach. The last part of the work includes a conclusions and bibliography.

2 Medical Images Database

It is necessary to have appropriate amount of real case data to test new developed as well as existing image analysis algorithms. Probably, the most popular database of FNB images and nuclei features is Wisconsin Database of Breast Cancer (WDBC). However, the quality of images delivered in the set is unsatisfactory for image analysis methods described in the paper. Because of that we decided to use our own data set.

The database contains 500 images of the cytological material obtained by FNB. The material was collected from 50 patients of outpatient clinic ONKOMED in Zielona Góra. It gives 10 images per case which was recommended amount by specialists from the Regional Hospital in Zielona Góra [6, 11]. This number of images per single case allows correct diagnosis by a pathologist. The set contains 25 benign and 25 malignant lesions cases. Biopsy without aspiration was performed under the control of ultrasonograph with a 0.5 mm diameter needle. Smears from the material were fixed in spray fixative (Cellfix of Shandon company) and dyed with hematoxylin and eosin (h+e). The time between preparation of smears and their preserving in fixative never exceeded three seconds. The images were recorded by SONY CDD IRIS color video camera mounted atop an AXIOPHOT microscope. The slides were projected into the camera with 40x and 160x objective and a 2.5x ocular. One image was generated for enlargement 100x and nine for enlargement 400x. Images are BMP files, 704x578 pixels, 8 bit/channel RGB. All cancers were histologically confirmed and all patients with benign disease were either biopsied or followed for a year.

3 Segmentation of the Nuclei

Classification of tumor malignancy requires isolating nuclei from the rest of the image. In literature, many different approaches have been already proposed to extract cells from microscope images [1, 4, 5, 6, 8, 9, 10, 11, 14]. This task is usually done automatically, using one of the well known methods of image segmentation [7, 13, 16, 18]. Unfortunately, reliable cell segmentation is a challenging task. Very often cells cluster and overlap together and their boundaries are blurred. Moreover, attempts to generalize segmentation approaches proposed in literature usually fail because such methods work correctly only for specific images. Slides from various sources may vary significantly depending on the method of smear preparation. In order to deal with these problems, we have developed automatic segmentation procedure that integrates results of image segmentation from two different methods. Proposed algorithm uses adaptive thresholding segmentation to distinguish all dark objects (nuclei, red blood cells and others) from bright background. Next, the color information of found objects is employed to classify them using Fuzzy C-Means clustering algorithm. This step is performed to eliminate objects which are not nuclei and deliver no important diagnostic information.

The key idea of thresholding is to separate objects from background based on pixel intensity fluctuations. There are many approaches for computing the optimal threshold [7, 13, 16, 17]. Most of them are based on histogram analysis or examination of pixel intensity. Such global thresholding methods work correctly for images with uniform illumination, but they fail on images with strong illumination differences. Unfortunately, effects of non-uniform lighting conditions can be observed on most microscope images used in this work. In order to solve the problem adaptive threshold method was applied. Local threshold is calculated for each pixel using intensities of pixels from its neighborhood. This area was defined as a square window

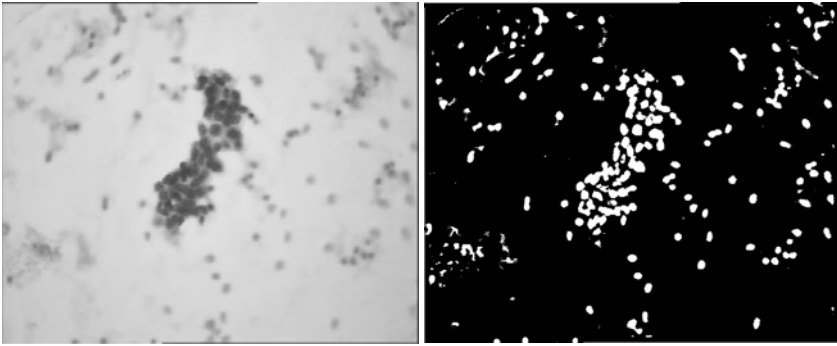


Fig. 1 Sample thresholding segmentation: original image (left), adaptive thresholding (right)

of size 20×20 pixels and the threshold is a mean intensity value of pixels inside the window. Such an approach eliminates the problem with non-uniform illumination but causes that insignificant objects with high local contrast (eg. red blood cells) are also separated from the background (Fig. 1).

In order to filter the objects which are useless for malignancy classification procedure, fuzzy clustering algorithm was applied [2, 10, 13]. The idea of the image segmentation using clustering algorithms boils down to a search for clusters of pixels in color space. Each cluster represents objects that are characterized by similar color. In the considered case Fuzzy C-Means (FCM) algorithm was applied to calculate centers for 3 clusters and to compute clusters membership degrees of pixels in RGB space. The clustering procedure of FCM algorithm is based on finding the local minimum of the nonlinear cost function given by equation 1 using the Picard iteration through first order conditions for stationary points.

$$J(\mathbf{U}, \mathbf{V}) = \sum_{x=1}^X \sum_{y=1}^Y \sum_{k=1}^C \mu_{x,y,k}^m f_c(\mathbf{c}_{x,y}, \mathbf{v}_k), \quad (1)$$

where the matrix $\mathbf{U} \in \mathbb{R}^{X \times Y \times C}$ contains the membership degrees of pixels to the defined clusters, $\mathbf{v}_k \in \mathbb{R}^q$ is a vector of the coordinates of the k -th cluster center, X and Y defines the size of the analyzed image, f_c is a function used to determine the distance between the data points and cluster centers, $\mu_{x,y,k}$ is the membership of the (x,y) -th pixel in the fuzzy cluster k , $m \in (1, \infty)$ is the fuzziness of the clustering procedure, $\mathbf{c}_{x,y}$ is a vector of (x,y) -th pixel parameters which describes its color. The function f_c defines the metric used to determine the distance between the data points and cluster centers:

$$f_c(\mathbf{c}_{x,y}, \mathbf{v}_k) = \|\mathbf{c}_{x,y} - \mathbf{v}_k\|^2 = (\mathbf{c}_{x,y} - \mathbf{v}_k)^T \mathbf{A} (\mathbf{c}_{x,y} - \mathbf{v}_k). \quad (2)$$

The matrix \mathbf{A} from the expression (2) is used to tune the shape and orientation of the clusters in space. In the simplest approach, the matrix \mathbf{A} is unitary, thus the distance measure $f_c(\mathbf{c}_{x,y}, \mathbf{v}_k)$ is an Euclidean norm. In this case, the study metric is defined

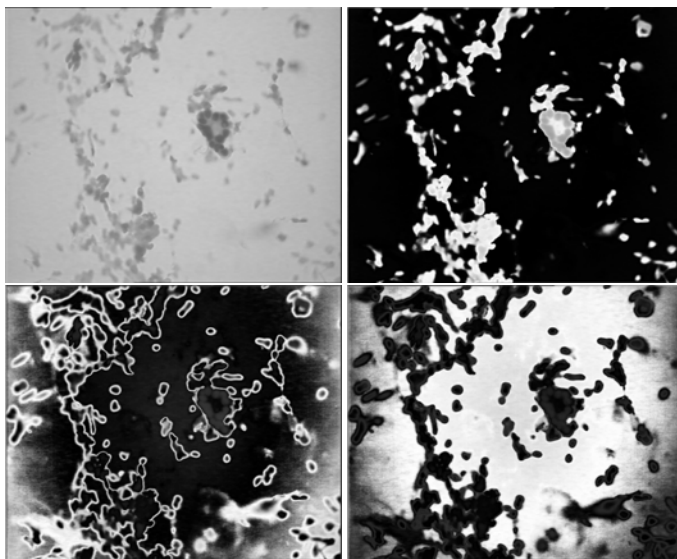


Fig. 2 Wrong FCM segmentation: original image and membership degrees for cluster 1,2, and 3 (brighter color indicates higher membership degrees)

as a Euclidean distance which measures color dissimilarity in RGB color space. Detailed expressions for iterative updating the cluster centers and membership values can be found in the following paper [2].

Membership values computed by FCM clustering are then used during segmentation to separate nuclei from background. This task is done by thresholding membership values of nuclei cluster. Threshold value is chosen arbitrary from the range 0.6-0.8. The cluster describing nuclei is detected by choosing the cluster with the lowest mean of RGB values of its center. Most of the considered images are correctly segmented using described procedure. However, some images in which only few nuclei are present are problematic for FCM based segmentation. Problem arises when clustering algorithm is not able to generate correct cluster that describe nuclei due to small number of pixels representing the nuclei respectively to the whole image (Fig. 2). Some modifications of FCM based segmentation were proposed to solve this problem. The problem of poor representation of nuclei can be circumvented by using group of randomly chosen images to determine the clusters centers. Then, such data is used to compute membership values for each image. Segmentation results for such procedure are correct even for images in which only few nuclei are present.

FCM-based segmentation is able to correctly distinguish nuclei from background and red blood cells but behaves worse for clustered nuclei. Much better results for separation of clustered nuclei are obtained using adaptive thresholding segmentation but in this case thresholding is not able to distinguish nuclei from red blood cells. In order to combine the advantages of both methods, results of both segmentation algorithms are fused together by multiplying segmented images.

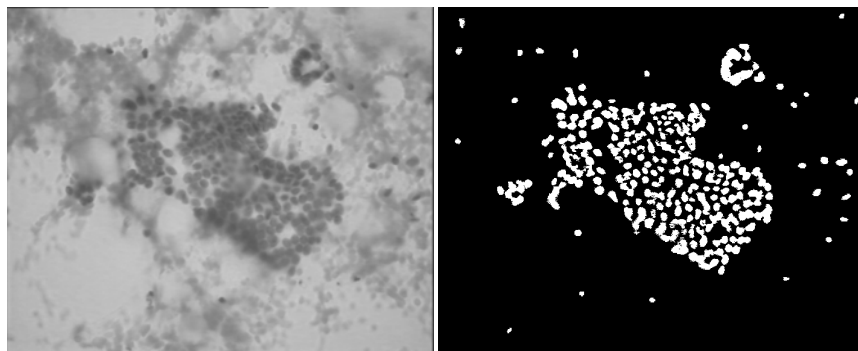


Fig. 3 Sample results for FCM and adaptive threshold segmentation: original image (left), extracted nuclei (right)

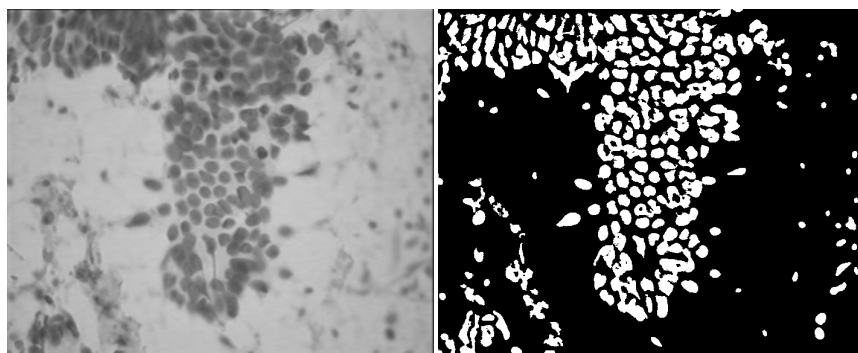


Fig. 4 Sample results for FCM and adaptive threshold segmentation: original image (left), extracted nuclei (right)

Segmentation results obtained for developed procedure of nuclei extraction allowed classify the degree of breast cancer malignancy with relatively high accuracy. However, it must be noted that proposed method usually will not be able to separate strongly clustered nuclei with blurred boundaries and this problem may degraded the results of the classification. Sample segmentation result are presented in Fig. 3 and 4.

4 Experimental Results

The FNB images were segmented with aforementioned method. Then a set of features was extracted from the result images. Finally classification was performed. In order to measure the effectiveness of the segmentation algorithm the same procedure of extracting and classifying was applied for manual segmentation.

For each cell following features were extracted: area, perimeter, eccentricity, major axis length, minor axis length, luminance mean, luminance variance and distance

from the centroid of all nuclei on the image. For each slide mean and variance of certain features were computed. Finally all input variables were normalized.

It was decided to use four well known classification algorithms such as k-nearest neighbor (with $k = 9$), naive Bayes classifier (with normal kernel distribution), decision trees (with GINI criterion) and classifiers ensemble [3, 12]. However, it must be mentioned that ensemble of classifiers is not a separate classification technique and its classification procedure is based on the results of others classifiers used in the experiments. Simply, the answer of classifiers ensemble is determined by voting procedure and class that gathers majority vote wins and represents the answer of the classifiers ensemble.

The prospective accuracy of the resulting classifiers was tested using the leave-one-out validation technique [12]. Since the number of samples is relatively small, using chosen classification algorithms with leave-one-out is computationally tractable and allows for accurate estimation of the error.

Analyzing segmented images one might see that the quality of the segmentation is not perfect. Many cells are joined or the centers of the cells are removed. However, classification using features such as area, luminance statistics or general distribution of cells on image gave very promising results. The discriminative power of individual features was estimated with previously indicated classifiers and results in form of recognition rates are presented in Table 1. Recognition rate is defined

Table 1 Result classification rate for particular features using automatic segmentation method (in bracket results for manual segmentation)

Feature	Statistic	kNN	Naive Bayes	Decision trees	Ensemble classifiers
area	mean	59.0 (81.9)	62.4 (84.7)	61.5 (82.5)	59.8 (83.9)
	variance	68.1 (86.4)	71.2 (84.4)	69.5 (82.8)	71.5 (86.4)
perimeter	mean	68.9 (84.4)	71.5 (84.4)	67.2 (81.4)	69.5 (83.6)
	variance	76.6 (83.3)	77.5 (83.3)	76.6 (80.0)	76.9 (83.3)
eccentricity	mean	65.8 (52.5)	70.1 (56.1)	62.7 (50.6)	67.2 (50.0)
	variance	68.9 (52.8)	73.5 (57.2)	66.7 (54.4)	68.9 (54.7)
major axis length	mean	61.5 (81.4)	68.9 (83.6)	60.4 (80.3)	62.7 (81.9)
	variance	67.8 (79.7)	69.2 (80.6)	68.7 (77.2)	69.8 (78.9)
minor axis length	mean	59.0 (83.9)	65.5 (83.6)	56.7 (79.4)	60.7 (83.6)
	variance	67.8 (83.9)	65.0 (83.9)	68.4 (83.6)	65.8 (84.2)
luminance mean	mean	71.2 (67.8)	73.5 (67.5)	68.9 (66.1)	72.9 (69.2)
	variance	63.2 (64.1)	65.8 (67.5)	61.5 (61.4)	64.4 (65.8)
luminance variance	mean	71.8 (54.2)	71.8 (56.4)	72.9 (54.2)	72.9 (56.9)
	variance	68.9 (55.0)	71.5 (54.7)	65.0 (56.4)	68.1 (56.4)
distance from centroid	mean	74.4 (78.3)	78.3 (79.4)	70.9 (76.7)	74.6 (76.7)
	variance	66.4 (75.3)	70.1 (76.7)	67.8 (70.8)	69.2 (75.0)

Table 2 Result classification rate for sets of features using automatic segmentation method (in bracket results for manual segmentation)

Input variables feature (statistic)	kNN	Naive Bayes	Decision trees	Ensemble classifiers
area (mean), area (variance), perimeter (mean), luminance mean (mean), luminance variance (mean), major axis length (variance), minor axis length (variance), distance from centroid (mean), distance from centroid (variance)	84.0 (93.7)	82.6 (92.3)	82.6 (90.3)	84.0 (93.4)
area (mean), perimeter (mean), perimeter (variance), luminance variance (mean), distance from centroid (mean)	80.3 (91.5)	79.5 (91.5)	79.2 (91.7)	81.8 (93.4)
area (mean), luminance mean (mean), major axis length (variance), distance from centroid (variance)	82.1 (93.2)	79.5 (92.9)	80.6 (89.7)	81.8 (93.4)
perimeter (mean), perimeter (variance), luminance mean (mean), distance from centroid (mean)	82.9 (91.5)	83.8 (92.3)	81.5 (90.0)	84.9 (91.7)

as percentage of successfully recognized cases to the total number of all cases. The results obtained with aforementioned segmentation method was compared with the manual segmentation results. Some features, specially describing shape of cells such as area or major axis length were significantly worse for automatic segmentation. The reason is low quality of the segmentation process. However, some other features, mainly characterizing luminance and texture, mean luminance for instance, gave better recognition rate for automatic than manual segmentation.

To find a set of features the best discriminative benign and malignant cases sequential forward selection was applied. Taking into account the fact that different subsets can be optimal for different classifiers, two approaches was applied to forward selection. First consider the same subset of features for each classifier and classifiers ensemble was used to assess the final quality of subset, and second approach assumes that the classifiers have different optimal subsets of features. Comparison

of best subsets of the features for each specific classification method is presented in Table 2.

The classification rate for automatic segmentation is about 10% worse than for manual one and amounts around 80%. Such results are very promising according to the fact that the quality of the segmentation method might be considerably improved by detecting overlapping and joined objects. Also, it must be mentioned that in the experiment each image was treated as a separate case. The best classification rate obtained was 84.9%.

5 Conclusions

The main objective of the described work was to develop an effective segmentation method for breast cancer malignancy classification problem. The accuracy of the segmentation process obtained in developed approaches is very promising, despite the fact that current methods are not able to handle properly overlapped nuclei yet. The segmentation quality should be improved by detecting overlapped cells on images and splitting or excluding them from the classification process. Also, recognition rate should be improved by adding more sophisticated features not tested during current investigations. Another challenge will be applying the whole segmentation and classification system for virtual slides generated by virtual scopes which are able to produce images with the resolution of 50000x50000 or even higher. Such huge slides require completely new way of analysis. It will be crucial to develop algorithms to find interesting parts of the slides containing valuable information for further processing and diagnosis.

References

- [1] Al-Kofahi, Y., Lassoued, W., Lee, W., Roysam, B.: Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images. *IEEE Trans. on Biomedical Engineering* 57(4), 841–852 (2010)
- [2] Bezdek, D.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)
- [3] Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: *Classification and Regression Trees*. Chapman & Hall, Boca Raton (1993)
- [4] Clocksin, W.F.: Automatic segmentation of overlapping nuclei with high background variation using robust estimation and flexible contour models. In: *Proc. 12th Int. Conf. Image Analysis and Processing (ICIAP 2003)*, pp. 682–687 (2003)
- [5] Cloppet, F., Boucher, A.: Segmentation of overlapping/aggregating nuclei cells in biological images. In: *Proc. ICPR*, pp. 1–4. IEEE, Los Alamitos (2008)
- [6] Filipczuk, P., Kowal, M., Marciniak, A.: Feature selection for breast cancer malignancy classification problem. *J. Medical Informatics & Technologies* (15), 193–199 (2010)
- [7] Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Prentice Hall, New Jersey (2001)

- [8] Hrebień, M., Steć, P., Obuchowicz, A., Nieczkowski, T.: Segmentation of breast cancer fine needle biopsy cytological images. *Int. J. Appl. Math and Comp. Sci.* 18(2), 159–170 (2008)
- [9] Jeleń, Ł., Fevens, T., Krzyżak, A.: Classification of breast cancer malignancy using cytological images of fine needle aspiration biopsies. *Int. J. Appl. Math. and Comp. Sci.* 18(1), 75–83 (2008)
- [10] Kowal, M., Korbicz, J.: Segmentation of breast cancer fine needle biopsy cytological images using fuzzy clustering. In: Kornacki, J., Raï, Z., Wierzchoń, S., Kacprzyk, J. (eds.) *Advances in Machine Learning I*, pp. 405–417. Springer, Heidelberg (2010)
- [11] Marciniak, A., Obuchowicz, A., Monczak, R., Kołodziński, M.: Cytomorphometry of Fine Needle Biopsy Material from the Breast Cancer. In: *Proc. 4th Int. Conf. on Computer Recognition Systems (CORES 2005)*, pp. 603–609. Springer, Heidelberg (2005)
- [12] Mitchell, T.M.: *Machine Learning*. McGraw-Hill, New York (1997)
- [13] Naz, S., Majeed, H., Irshad, H.: Image Segmentation using Fuzzy Clustering: A Survey. In: *Proc 6th Int. Conf. Emerging Technologies (ICET 2010)*, pp. 181–186 (2010)
- [14] Obuchowicz, A., Hrebień, M., Nieczkowski, T., Marciniak, A.: Computational intelligence techniques in image segmentation for cytopathology. In: Smolinski, T.G., Milanova, M.G., Hassanien, A.-E. (eds.) *Computational Intelligence in Biomedicine and Bioinformatics. Studies in Computational Intelligence*, vol. 151, pp. 169–199. Springer, Heidelberg (2008)
- [15] Otsu, N.: A Threshold Selection Method from Gray-level Histograms. *IEEE Trans. Sys. Man. and Cyber.* (9), 62–66 (1979)
- [16] Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. *J. Electronic Imaging* 13(1), 146–165 (2003)
- [17] Shapiro, L.G., Stockman, G.C.: *Computer Vision*. Prentice Hall, New Jersey (2002)
- [18] Suri, J.S., Setarhdan, K., Singh, S.: *Advanced Algorithmic Approaches to Medical Image Segmentation*. Springer, London (2002)
- [19] Underwood, J.C.E.: *Introduction to biopsy interpretation and surgical pathology*. Springer, London (1987)
- [20] Wolberg, W.H., Street, W.N., Mangasarian: Breast cytology diagnosis via digital image analysis. *Analytical and Quantitative Cytology and Histology* 15, 396–404 (1993)

Application of Pattern Recognition Techniques for the Analysis of Histopathological Images

Adam Krzyżak, Thomas Fevens, Mehdi Habibzadeh, and Łukasz Jeleń

Abstract. In this paper we discuss applications of pattern recognition and image processing to automatic processing and analysis of histopathological images. We focus on two applications: counting of red and white blood cells using microscopic images of blood smear samples and breast cancer malignancy grading from slides of fine needle aspiration biopsies. We provide literature survey and point out new challenges.

Keywords: CBC, microscopic medical images denoising, binarization, segmentation, edge preservation, granulometry, fine needle aspirates, breast cancer malignancy grading.

1 Introduction

Automatic detection of pathologies from histopathological images is currently very active and important area of research. In the present paper we will survey two applications of pattern recognition and image processing in this emerging field: automatic processing of blood smear images and automatic grading of breast cancer fine needle biopsy slides. The paper is organized as follows: in sections 2-5 we will review processing of blood smears and in sections 6-9 we will focus on cancer grading.

Adam Krzyżak · Thomas Fevens · Mehdi Habibzadeh
Department of Computer Science and Software Engineering, Concordia University,
1455 De Maisonneuve Blvd. West, Montréal, Québec, Canada H3G 1M8
e-mail: [{{krzyzak, fevens, me_hab}@encs.concordia.ca}](mailto:{krzyzak, fevens, me_hab}@encs.concordia.ca)

Łukasz Jeleń
Faculty of Life Science and Technology, Wrocław University of Environmental and
Life Science, Norwida 2527, 50-375 Wrocław, Poland
e-mail: lukasz.jelen@up.wroc.pl

2 Manual Analysis of Blood Smear Images

Analysis of microscopic medical images is an important interdisciplinary problem involving both physicians and computer scientists. One of the important and active areas of research is the problem of counting blood cells (CBC) [1, 2, 3] which is used as screening test to check such disorders as infections, allergies, problems with clotting, and it helps diagnosing and managing a large number of diseases. In practice a panel of tests is carried out that examine different blood components such as counting white blood cells (WBC) [2, 4], white blood cells differential, counting red blood cells (RBC) [2], checking for signs of disease and the counting the number of infected cells. Blood cell counting and blood film examination are widespread diagnostic techniques [3]. A blood smear is obtained by drawing blood from a vein and placing a drop on a glass slide [3]. The blood film is stained [3] using e. g., Wright's, Giemsa, or May-Grunwald staining techniques and imaged with a transmission light microscope. The definitive diagnosis of blood smear infection is done by manually finding disorders and abnormalities in blood films through a microscope, counting blood smear particles and cells with disorders, which are not only a time consuming task but also prone to human error. The erythrocytes and leukocyte types that the current equipment is able to manage are restricted to few classes [5] and staining process requires expensive chemicals.

As mentioned, the microscope inspection of blood slides provides important qualitative and quantitative information concerning the presence of hematic pathologies [6], however the number of different sub-cell types that can come out especially for WBC count is relatively large and typically more than 20 [5]. Normal peripheral blood contains the following types of leukocytes (the numbers in brackets give the typical proportion of the cell type): segmented neutrophil (40- 75%), lymphocyte (25-33%), monocyte (2-8%), eosinophil granulocyte (1-4%), band neutrophil (1-3%), plasma cell (0.2-2.8%), basophil granulocyte (0.5%), and atypical lymphocyte. Other cell types which are observed in certain diseases include: metamyelocyte, myelocyte, promyelocyte, myeloblast and erythroblast [3] and this increases the difficulty in building a feasible system. This process can be automated by computerized techniques which are more reliable and economic. Therefore there is always a need for the development of systems to provide assistance to hematologists and to relieve the physician of drudgery or repetitive work. So, more systems for automatic processing of medical images are being developed and during blood film examination, the individual types of blood smear particles (leukocytes and erythrocytes) are enumerated yielding so called differential count.

Our goal is to develop and validate the necessary image and pattern recognition processing algorithms to quantify and detect microscopic particles on slides to enhance automated system to characterize blood health status of patient. In essence that will enable us to determine the fast, accurate mechanism of segmentation and gather information about distribution of microscopic particles which are help to diagnose status of abnormality or normality and represent a factor of combatively and quality for the modern laboratories of clinical analysis.

3 Automatic Processing of Blood Smear Images

During blood film examination, the individual types of blood smear particles (leukocytes and erythrocytes) are enumerated and then blood films are usually made to investigate hematological problems [1, 2]. The history of research into automated blood slide examination dates back to 1975, see Bentley & Lewis [7]. However it is only recently that digital photography, computer speed, RAM size and secondary storage capacity have made automatic blood processing possible. The analysis of blood slides must be fully automated to be useful [8]. Due to complexity of the problem at hand (Costin et al. [9]) most of the papers are limited to image-based comparisons based on red cells segmented either manually, see Bentley & Lewis [7], Albertini et al. [10], or semi-automatically, see Robinson et al. [11], Costin et al. [9] and Gering & Atkinson [12].

There is a vast amount of literature dedicated to differential blood counts. A key step in automating this process is the segmentation of cell boundaries. Initial success on segmentation of medical images was obtained with graph theory (Martelli [13], Fleagle et al. [14], Fleagle et al. [15]) which was used to navigate around edge pixels found in an image. However this approach has involved images of single objects manually located in an image, and does not address the problems of multiple objects in the image, object location, removal of extraneous edges (internal to the cell), or the selection of suitable starting and ending points for the graph search. Furthermore, thresholding has been used to pre-process images as an aid to segmentation (Gonzalez & Woods [16]). With red blood cell images this causes problems due to the pale nature of the interior of the cells, which then necessitates further processing. Adjouadi and Fernandez [17] find the cell borders using eight-directional scanning within thresholded images of normal blood. The problem with this approach is that it would not find the whole of the border of severely deformed cells as these contain edge points that would not be reached by any of the eight scan-lines. Moreover the process does not result in identification of the points within the contours. Di Ruberto et al. [18] follow thresholding with a segmentation using morphological operators combined with the watershed algorithm [19]. However their work is aimed at segmentation of red blood cells containing parasites and is designed to increase the compact nature and roundness of the cells. Such assumption of roundness is not appropriate for segmentation of all particles for the purpose of classification, because blood smears may contain deformed red blood cells or some WBC types [1] which are not 100% convex and circular. The method is also complicated requiring nine intermediate steps and does not result in border identification. It also requires some preprocessing which is not always applicable for all possible slides and then there is no dynamic way to better control image acquisition.

Another popular automatic approach to border detection is that of active contours, or snakes (Kass, Witkin & Terzopoulos [20]), which can be applied either to the original image or to an edge image. However when used to identify cell borders, the resulting contours do not correspond with the exact borders of the cells (Ongun et al. [21], Wang, He & Wee [22]), which would cause problems with subsequent RBC classification, where the exact boundary shape is important. Other problems

with the use of contours for images of peripheral blood smear slides include the initial positioning of the multiple contours required; the tendency of the contours to find the inner pale cell areas in addition or instead of the outer edges; and the failure of contours to identify pixels interior to the contour. Other works using active contours for tracking boundaries of WBC include Ongun [23] but it could not deal with WBC overlapping problem. Lezoray [24] proposed a region-based WBC segmentation strategy using seed flooding. However, it relied greatly on the proper seed extraction using prior knowledge of color information. Kumar [25] defined a new edge operator and tried to get precise nucleus edge. But it required relatively weak edge existing between red blood cell (RBC) and background, which was often missing. An automated system where cells are segmented using active contour models (snakes and balloons, initialized by morphological operators) are presented in [26]. The shape and texture features are used for classification. A two step segmentation process is used by Sinha and Ramakrishnan [27]. First the HSV transformed image is clustered using k-means followed by an EM-algorithm. The shape, color, and texture features are then used in a neural network classifier. A mean-shift-based color segmentation procedure applied to leukocyte images is described in [28]. Segmentation is performed in the $L^*u^*v^*$ color space. A watershed-based segmentation is used in [29]. First a sub-image containing a leukocyte is separated from the cell image. The nucleus region is then detected by scale-space filtering and the cytoplasm region by watershed clustering of the 3-D data. WBC classification in recent work Hamghalam et al. [30] utilizes Otsu's thresholding method to segment nuclei. The results are independent of the intensity differences in Giemsa-stained images of peripheral blood smear and active contours are used to extract precise boundary of cytoplasm.

As mentioned previously, the nature of microscopic particles is not simple and automatic processing of images in medicine is a complicated task. This is because some of the basic tasks to be performed such as pre-processing, segmentation, classification, object recognition and inference require extensive understanding of the specific problem. This requires comprehensive knowledge in many disciplines such as medicine, computer science, image and signal processing.

4 Methodology and Algorithms

In the next few sections we will review basic steps for processing microscopic blood smear images.

4.1 Image Acquisition and Denoising

The first step is to convert RGB channels images to the green channel as it is more reliable than the red or blue channels for noisy and distorted images. The next step is choosing an effective denoising tool. To design a reliable automated segmentation system that may be used under different conditions such as a variety of microscopic staining techniques, types of chemical materials used, microscope types,

illumination conditions, human error, etc., a pre-processing step is required. The accuracy of this stage affects the system performance. There are wide variety techniques for enhancing image quality. One of the most practical and widely used denoising technique is wavelet shrinkage approach which thresholds the wavelet coefficients of an image. Wavelet coefficients having small absolute value are considered to encode mostly noise and very fine details of the signal. In contrast, the important information is encoded by the coefficients having large absolute values. Removing the small coefficients and then reconstructing the signal could produce signal with lesser amount of noise. The biggest challenge in the wavelet shrinkage approach is finding an appropriate threshold value [31].

The wavelet shrinkage approach can be summarized as follows:

1. Apply the wavelet transform to the signal.
2. Estimate a threshold value.
3. Remove (zero out) the coefficients that are smaller than the threshold.
4. Reconstruct the signal (apply the inverse wavelet transform)

In [32, 33] Daubechies wavelet with soft thresholding and Bivariate Shrink together with PSNR ratio has been used. In using soft thresholding based on following concepts the user should calibrate the parameters of the algorithm. The optimal thresholding obtained by using soft thresholding which depends on experience and on the type of images.

In Figs. 1 and 2 we illustrate wavelet denoising including two simulation studies: one for images corrupted by moderate additive normal noise with deviation 30 and the second for highly corrupted by additive normal noise with deviation 100.

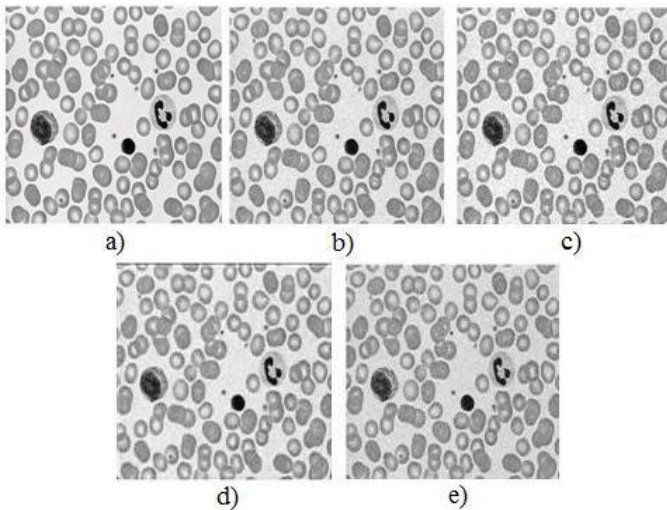


Fig. 1 a) original image (red channel); b) noisy image; c) median denoising; d) soft thresholding denoising; e) Bivariate denoising.

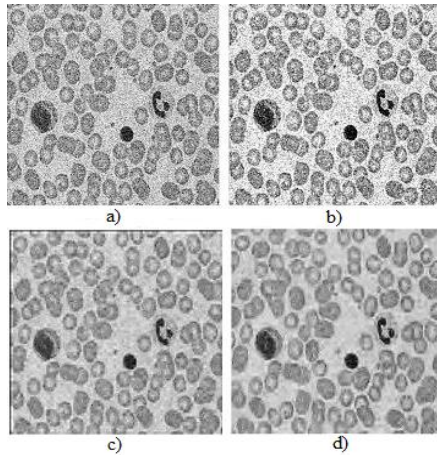


Fig. 2 a) noisy image; b) median denoising; c) soft thresholding denoising; d) Bivariate denoising.

Table 1 PSNR levels for various denoising techniques for images with moderate and high noise.

		Additive Noise deviation	
		30	100
PSNR	Noisy Image	19.2149	10.4516
	Denoised Image using Median	25.5666	16.5183
	Denoised Image using Thresholding	23.5460	18.3421
	Denoised Image using Bivariate	27.6236	20.4822

For moderate noise and high noise, the PSNR experimental results are summarized in Table 1. From the experimental results it can be concluded that for moderate noise the Bivariate Shrink filter produces best results. It produces the maximum PSNR for the output image compared to the other filters. However, Bivariate output, is somehow blurred and some post-processing involving de-blurring and edge preserving may be needed. For images heavily corrupted by noise with low PSNR value (10.4516) the Bivariate Shrink filter is again best. It produces the maximum and acceptable PSNR for the output image compared to the other filters. It can also be observed that for high noise levels soft thresholding produces better results than the classical median filter.

4.2 Edge Preservation

The aim of the next edge preservation step is to recover degraded and blurred images while reducing the negative effects of noise such as blurred edges produced by the Bivariate Shrink filter. This step can serve as preliminary step prior to binarization

and object segmentation. Different edge preservation techniques have been used in practice. They include median filter [34], symmetrical nearest neighbor (SNN) filter [35], convolution kernel filters [36], preserving color reduction method [37], bilateral techniques [38], and the Kuwahara filter [39]. In computer simulations we have learned that Kuwahara filter works best. This can be justified by intrinsic characteristics of microscopic particles for which Kuwahara filter yields the sharpest edges which leads to better binarization in next step (see Fig. 3). However, the output may be somewhat toothy and jagged.

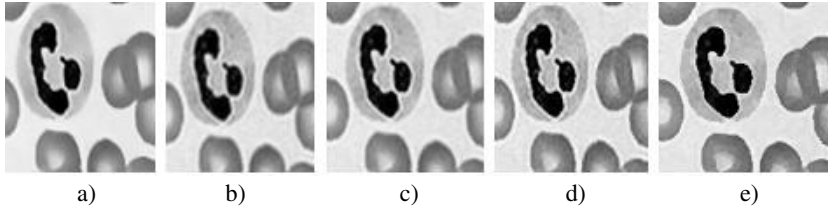


Fig. 3 a) edge preservation with Bilateral, b) convolution kernel, c) EDGEPS [37], d) SNN, e) Kuwahara filters, respectively.

4.3 Binarization

After denoising and edge enhancement, binarization is the third step which allows to extract some features, having sub images and get ready to apply new technique for different purpose over the images. Generally, binarization methods can be applied with global and local thresholding. Different binarization methods include the approaches of Niblack [40], Bernsen [40], Sauvola [41] and Otsu [42]. Computer experiments with different samples and initial conditions (see Fig. 4) show that Niblack approach is the most reliable method to maintain disjoint components which is crucial in avoiding over or under segmentation.

Mehdi et al [43] proposed a modified binarization method that merges Niblack and Otsu approaches. This process reduces limitations and drawbacks of each of them. Niblack uses local thresholding based on average and standard deviation of a local area. The size of the window must be large enough to suppress the noise but at the same time small enough to preserve local details. In practice, a window

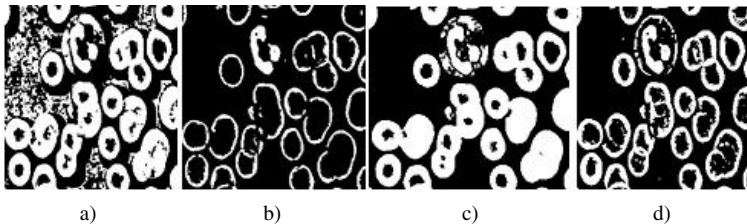


Fig. 4 Binarization methods: a) Bernsen; b) Sauvola; c) Otsu; and d) Niblack

size of 15×15 works well in all available image databases. The Niblack method tends to result in overlapping objects that are too close to one another which in turn leads to false segmentation results. In the modified version, pixels are labeled as background pixels if they are labeled as either background pixels in Niblack or in Otsu and the remaining pixels are kept as foreground pixels (objects). Using this merging process, we mitigate the problem of extra small spurious regions produced by the Niblack algorithm.

4.4 Size Estimation

Binarization and some post-processing to enhance the quality of binary image is followed by feature extraction which helps to differentiate various types of particles in the image. A normal blood cell is one of two major particles: a RBC with a normal probability distribution function (PDF) with average size around $6.0\text{--}8.5 \mu\text{m}$ or a WBC with average size around $7\text{--}18 \mu\text{m}$ which includes a nucleus and cytoplasm is about 1-3 times bigger than normal and mature RBCs. We use size characteristics as an effective factor to distinguish between the two main types of cells. Granulometry [44] can determine the size distribution of image objects without explicitly segmenting each object first. According to normal blood PDF and RBC to WBC ratio, the maximum regional peak in pattern spectrum diagram correlates to the number of RBCs with an acceptable RBC radius size. We summarized the granulometry algorithm in the next section.

Granulometry Algorithm

Granulometry is concerned with size distribution of cells in binary images. It uses structure elements which are morphologically dilated to the maximum size and applied to the image. The shape of structure element depends on the type of objects under processing. During the process granulometric density function is determined [7, 8].

Granulometric algorithm starts by applying opening morphology along with defined structure element (SE). In normal blood smear images, all available particles are approximately circular. Hence, we select (disk) shape as default and basic structure element for granulometric algorithm. In ideal output, we expect only one peak for a single complete circle, but the incomplete circular object shown in Fig. 5 produce local maxima. We call this undesirable effect an edge fracture. We just observe that after applying the edge detection and skeletonisation algorithms to real cell images which are typically not complete curves the observed circular pieces are regarded as a new objects surrounded between two ideal complete circles. Consequently we can expect in granulometric output at least two local regional peaks. By this simple work, we find that blood smear particles are not complete circular object and there are always discrete components on curve tracer, which is another reason for undesirable local maxima.

Overall, applying granulometry to RBCs images in normal blood smear can be very reliable in determination and estimating their size. But for abnormal samples

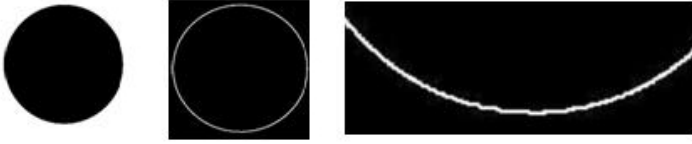


Fig. 5 Granulometry over simple circle

with different shapes or with extra overlapping between the particles granulometric approach may fail.

4.5 Segmentation

There are different methods which are directly or indirectly subjected to separation and segmentation objects in disjoint images such as active contours and watershed. Typically watershed is incorporated into the immersion and toboggan methods [45]. The accuracy and efficiency of watershed segmentation over images is directly related to the previous steps. In practice, watershed algorithm works best for smooth convex objects that don't overlap too much. It cannot be an efficient approach in all microscopic images with extra overlapping which can happen for some diseases.

5 Conclusion: Segmentation of Microscopic Imagery

In this paper, a simple and step-by-step efficient algorithm has been presented together fully automated detection and segmentation of microscopic imagery. Experimental results indicate that the current analysis is accurate and offers remarkable segmentation accuracy. The performance of the proposed method has been evaluated by comparing the automatically extracted particles with manual segmentations and other traditional techniques [43]. Furthermore, the introduced method being simple and easy to implement is best suited for biomedical applications in clinical settings.

6 Cancer Cells Grading

Automatic cancer grading is a very challenging task due to large variation in cancer imaging and analysis. In the remainder of the paper we shall focus on automatic malignancy grading of breast cancer fine needle aspiration biopsies.

7 Breast Cancer Diagnosis

According to statistics breast cancer is one of the most deadly cancers among middle-aged women. Based on the data provided by the Breast Cancer Society of Canada about 415 women will be diagnosed with breast cancer each week in Canada. Most of the diagnosed cases can be fully recovered when diagnosed at an

early stage. Cancers in their early stages are vulnerable to treatment while cancers in their most advanced stages are usually almost impossible to treat. During the diagnosis process, the cancer is assigned a grade that is used to determine the appropriate treatment. Successful treatment is a key to reduce the high death rate. The most common diagnostic tools are a mammography and a fine needle aspiration biopsy (FNA). Mammography, which is a non-invasive method, is most often used for screening purposes rather than for precise diagnosis. It allows a physician to find possible locations of microcalcifications and other indicators in the breast tissue. When a suspicious region is found, the patient is sent to a pathologist for a more precise diagnosis. This is when the FNA is taken. A fine needle aspiration biopsy is an invasive method to extract a small sample of the questionable breast tissue that allows the pathologist to describe the type of the cancer in detail. Using this method pathologists can very adequately describe not only the type of the cancer but also its genealogy and malignancy. The determination of the malignancy is essential when predicting the progression of cancer.

In this section we will review the computerized breast cancer diagnosis, which is a very active field of research (see sec. 8). Additionally, we will also look over the less active field which is a computerized breast cancer grading (see sec. 9).

8 Computer–Aided Breast Cancer Diagnosis

Breast cancer diagnosis is a very wide field of research studying not only medical issues but also computer science issues. Breast cancer diagnosis is a multi-stage process that involves different diagnostic examinations.

Pattern classification is a well-known problem in the field of Artificial Intelligence concerned with the discrimination between classes of different objects [46]. We can use the same techniques in cancer diagnosis to assist doctors with their decisions. Cheng et al. [47] provided an extensive survey on automated approaches in mammograms classification and importance of computer assisted diagnosis. Since mammography is one of the preliminary tests performed to locate abnormalities in the breast tissue, it is used for screening purposes and has raised a lot of interest within the scientific community [47, 48, 49, 50, 51, 52, 53].

To the best of our knowledge, the computerized breast cytology classification problem was first investigated by Wolberg et al. in 1990 [54]. The authors described an application of a multi-surface pattern separation method to cancer diagnosis. The proposed algorithm was able to distinguish between a 169 malignant and 201 benign cases with 6.5% and 4.1% error rates, respectively depending on the size of the training set. When 50% of samples were used for training, the method returned a larger error. Using 67% of sample images reduced the error to 4.1%. The same authors introduced a widely used data-base of pre-extracted features of breast cancer nuclei obtained from fine needle aspiration biopsy images [55]. Later, in 1993, Street et al. [56] used an active contour algorithm, called 'snake' for precise nuclei shape representation. The authors also described 10 features of a nucleus used for

classification. They achieved a 97.3% classification rate using multi-surface method for classification.

The features described by the authors are mainly geometrical features of the nucleus. These features are:

- Radius – defined as an average of the radial line segments lengths from the centroid of the nuclei to the snake points on the boundary.
- Perimeter – is the length of the boundary of a polygon connecting snake points.
- Area – is a number of pixels inside the closed snake curve.
- Compactness = $\frac{\text{perimeter}^2}{\text{area}}$
- Smoothness of a nuclei contour – defined as an average difference between the length of a radial line and the mean length of the lines surrounding it.
- Concavity – a measure of nucleus concavity. This is performed by drawing chords between non-adjacent snake points and measuring the extent to which the boundary of the nucleus lies on the inside of each chord. The length of the chord that is outside of the nuclei is considered as a measure of the concavity. The concavity is larger when the length of the exterior chord increases.
- Concave points – measures number of concavities and not their magnitude.
- Symmetry – Here, the major axis (longest chord through the center) is found. Next, length difference between lines perpendicular to major axis to nucleus boundary in both directions are measured.
- Fractal Dimension of a cell – this is approximated using a 'coastline approximation' method. Authors measure the perimeter of the nucleus using increasingly larger segments. Next, they plot the obtained values on a log scale and calculate the downward slope which gives an approximation to the fractal dimension. Higher values of the feature provide higher probability of malignancy.
- Texture – authors define texture as an average gray scale intensity of the nucleus.

Based on the above features, Street [57], in his PhD Thesis, introduced a system called XCyt, that was later improved and described in 2000 [58]. In 1999, Lee and Street [59] described an iterative approach for automated nuclei segmentation as an addition to the previously described framework. In 2003, they introduced flexible templates to their iterative Generalized Hough Transform approach for segmentation. They created a set of predefined templates of a nuclei and each iteration shuffles the templates in such a way that those that were used the most often during the previous iteration are visited first to save time. The authors were able to segment nuclei with 78.19% accuracy [60]. They also introduced a neural network approach for classification stage, achieving 96% accuracy. Classification was based on the features previously described by Street et al. [56].

All work presented above was based on the Wisconsin Breast Cancer Database (WBCD) introduced by Mangasarian et al. [55]. This data-base consists of pre-extracted nuclear features and is widely used among researchers. Features included in the data-base are the features proposed by Street et al. [56]. WBCD [55] and its variations [61, 62] are the only data sets publicly available. Therefore, the majority of work in this field is performed on this data-base and involves research on different classification algorithms.

In 1998, Walker et al. [63, 64] introduced Evolved Neural Networks for breast cancer classification and tested their algorithm on WBCD data-base achieving 96% correctness. Nezafat et al. [65] used WBCD to compare several classification algorithms such as k-nearest neighbor classifier, radial-basis function, neural networks, multilayer perceptron and probabilistic neural networks. The authors showed that among these classifiers, multilayer perceptron with one hidden layer performed the most efficiently giving 2.1% error rate. Additionally they also compared and reported which of the features extracted by Wolberg et al. [54] were most significant for classification.

In 2002, Estevez et al. [66] introduced a different approach for classification based on the Fuzzy Finite State Machine, but their system performed rather poorly giving 19.4% error for the testing set of images. To extract features, the authors first manually segment nuclei from the image and then apply a low-pass filter and in the following step topological map of a nuclei is created. The extracted features are texture based. Motivation for them was that benign cell textures have bigger homogenous gray areas and more concentric contours than malignant cell textures.

Bagui et al. [67] recently introduced a classification algorithm applied to WBCD. The authors described a generalization of the rank nearest neighbor rule and obtained results that show a 97% recognition rate, which, according the authors, is better than that previously reported in the literature. From the above discussion we can deduct that majority of work in the field of breast cancer detection and classification was performed by Street et al. and Wolberg et al. We can find other approaches such as wavelet based approach of Weyn et al. [68]. Here the authors introduce a textural approach for chromatin description and claim that it has a 100% recognition rate.

Another approach is one introduced by Schnorrenberg et al. [69] that uses receptive fields for nuclei localization as an integral part of a bigger system, called 'BASS.' In 1996, they introduced a content-based approach [70] and provided an extensive survey on existing histopathological systems [71]. The authors presented two types of color-based features, luminance-based local features and global features. Luminance features were obtained from image RGB values. Global features are the variance and average of luminance in the image. They also introduce one texture measure that is calculated according to the luminance variance and current nucleus luminance. Approaches presented by Schnorrenberg et al. are mostly based on histological samples rather than cytological. In 2000, they presented a description of features used in their research [72] on classification of cryostat samples during intra-operative examination based on feed-forward neural networks achieving the highest accuracy of 76% on their own database.

In the literature we can also find some other approaches that involve segmentation of a breast cancer nuclei rather than classification. In 1996, Belhomme *et al.* [73] proposed a watershed based algorithm for segmentation of breast cancer cytological and histological images. Their algorithm is a more general version of the method described by Adams and Bischof [74]. The generalization involves the usage of numerous merging criteria. Authors use the segmentation principles described by Beucher in his PhD thesis [75]. This involves the decomposition of the segmentation

procedure into two steps. In the first step, the image is simplified based on a set of markers. The second stage involves region decomposition by the construction of the watershed lines [73]. The algorithm proposed by Belhomme *et al.* is the extension of the Beucher and Meyer [76] method by introduction of a general segmentation operator.

In 1998, Olivier *et al.* [77] introduced another extension to the watershed algorithm in addition to that of Belhomme *et al.* Their extension incorporates the color information in the image regardless of the color space. The authors compared their segmentation results against the segmentation performed by three experts and they reported the correctness of their method to be between 89.2% and 98.3% for the nuclei.

Another approach to nuclear segmentation is based on fuzzy *c*-means clustering and multiple active contours models described by Schüpp *et al.* [78]. The authors describe a level set active contours method, where the initial level set is obtained by the fuzzy *c*-means algorithm.

9 Computer-Aided Breast Cancer Grading

In the previous section we described different approaches for breast cancer diagnosis. Most of those systems discriminate only benign and malignant cases. For good diagnosis it is crucial to evaluate the malignancy grade. In cytology, the malignancy is graded according to the Bloom–Richardson scheme [79]. This system is based on grading of cells' polymorphy, the ability to reform histoformative structures, and mitotic index. All of these features are described by the Bloom-Richardson scheme as three factors that use a point based scale for assessing each feature. The malignancy of the tumor is assigned a grade that depends on the quantitative values of the above factors and is determined by the summation of all awarded points for each factor. Depending on the value, the tumor is assigned with low, intermediate or high malignancy grade.

In [80] we can see attempts at prognostication along with nuclear classification. For their grading approach, the authors used only nuclear features of a cell, which correspond to the second factor in Bloom-Richardson grading scheme. They were estimating the prognosis of the breast cancer according to these features. Further attempts for malignancy grading include VLSI approach introduced by Cheng *et al.* [81] in 1991 and applied in 1998 to breast cancer diagnosis [82]. In this method, the authors propose a parallel approach to tubule grading for histological slides. The authors divided their algorithm into four stages. The first stage consists of image enhancement for which purpose they use median filtering to remove artifacts. In stage two, the authors locate possible tubule formations by image thresholding with a threshold level known *a priori*. The next stage is a classification stage, where regions are classified as tubular formations. The features used in this study consists of brightness, bright homogeneity, circularity, size, and boundary colors. In the fourth stage, the authors count the number of tubular formations. The work presented by the authors not only deals with histology but also only mentions grading using only

one factor on the Bloom-Richardson scale. The authors showed time improvement of the parallel algorithm that grades tubules to $O(n)$ time while previously reported run time complexities were $O(n^2)$, where n is the size of the input data. In 1991, MacAulay *et al.* [83] introduced a graphics package for Bloom-Richardson grading of histological tissue. Their application acts as a typical graphics program that allows user to pick the nuclei from the image and perform some basic calculations. This process is almost completely user dependent. The authors provide an extensive description of the interface of the package but no further information on computation grading was found. Another approach found in literature is an algorithm based on wavelet texture description of chromatin [68]. This work was also performed on histological slides. The features calculated by the authors are calculated according to wavelet parameters and are divided into three groups. The first group are co-occurrence parameters that describe the color intensity in the image. The second set of parameters are densitometric parameters that are based on intensity values of the nucleus. The third group consists morphometric parameters that describe the geometry of the nucleus. Authors performed tests on their data-base of 83 histological slides and claim to have 100% classification rate. Such a high rate suggests a good separation between the classes.

In 2004, Gurevich and Murashov [84] proposed a method for chromatin structure analysis based on scale-space approach of Florack and Kuijper [85]. The authors claim that chromatin distribution corresponds to the grade of malignancy. This statement is supported by additional studies of Rodenacker [86, 87, 88] and Weyn *et al.* [89]. The authors also mention another approach to chromatin description. This method uses heterogeneity, clumpiness, margination and radius of particles and was introduced by Young *et al.* [90]. The algorithm of Guverich and Murashov uses topological properties of iso-intensity manifolds in the spatial extrema neighborhoods [84]. Their algorithm is able to measure the number of chromatin particles in the input image. For testing purposes the authors trained several classifiers achieving a classification rate between 72% and 85.4%. In 2006, Gurevich *et al.* [91] described a system for automatic analysis of cytological slides for the lymphatic system tumors. The authors used a Gaussian filter for segmentation of a nuclei from the previously extracted blue channel of the image. The feature extraction part of the proposed system is the same as in [84] plus an additional 47 features described by Churakova *et al.* [92]. These features include a well known and widely used morphological features such as the area of a nuclei, histogram features and features based on a Fourier spectrum of a nucleus [91]. In this paper, the same choice of classifiers was used as in [84] but the accuracy increased and is claimed by the authors to be above 90%. The authors did not provide an accurate error rate of their experiments and therefore it is difficult to assess the accuracy of the proposed system.

To the best of our knowledge, currently there is no publicly available database and most of the approaches presented in the literature are tested on the databases created by the authors, which makes the comparison of the obtained classification results with those reported in the literature difficult. The only commonly used database that we came across during this study is the Wisconsin Breast Cancer Database, which was described earlier in this thesis. This database is freely available from the

authors web page [54]. In this study, some of the proposed features are the same as in WBCD but the testing of the presented system on that database would be limited only to the classification stage due to the fact that WBCD is a database of pre-extracted features.

In 2005 a commercial system for automated histopathological tissue grading was released by QinetiQ [93]. According to the specifications and discussion with a pathologist, the results obtained by this system seem to be difficult to confirm. According to the authors, their system showed performance similar to the pathologists during clinical evaluation that was performed on 100 patients.

The most recent development in the field of automated breast cancer grading was described by Jeleń in his PhD thesis [94]. There are also other recent approaches by Naik *et al.* [95] and Jeleń *et al.* [96, 97, 98, 99, 100].

In [95] describe various segmentation methods such as level sets for classification of prostate and breast cancer histological slides. The described system was able to distinguish between low and high malignancy grades with 80.52% accuracy when automatic classification was used. The accuracy described by Jeleń in [94] was as high as 86.75% for cytological slides. The author in his thesis did an extensive study of the features and classification methods to determine a set of features and the classification method that will be able to classify the breast cancer malignancy into intermediate and high malignancy grades. Author also introduced a set of three new features that are used for the determination of the first factor of Bloom–Richardson scheme. These features were described in [96] and their discriminatory power were described in [98]. Features that were introduced by Jeleń include the area of grouped cells in the FNA slide (see Fig. 6), the number of groups that are visible on the slide and the third feature is a dispersion that describes if the cells in the image are grouped or dispersed. Beside a set of so called low magnification features author proposed the usage of 31 features that represented the nuclear structures of the cell. These features related to the second and third factor of the Bloom–Richardson grading scheme. In the thesis, the author performed a set of classification tests performed the calculations of the discriminatory power of the features to propose a set of features that are not correlated and provide the best classification results. From all of the tests, the author showed that the multilayer perceptron was the best performing classifier. The 34 element feature vector was reduced to 15 features. Fig. 7 shows graphically the correlation between the original set of 34 features. The features with the best discriminatory power were the three low magnification features described earlier and 12 nuclear features such as perimeter of a nucleus, convexity, x-centroid of the cell, nuclei orientation, its vertical projection, the ϕ_3 momentum feature, histogram mean, energy, textural homogeneity, red channel histogram mean, skew and width.

In [97] the authors did a comparative study of the discriminatory power of the low magnification features against the features based on the cell nucleus. From their study, one can notice that on average low magnification features perform better but the best classification was recorded for a feature vector that consisted of both types of features. In [99] the authors showed that the best classification was achieved for the multilayer perceptron when the fuzzy c-means segmentation was used. On

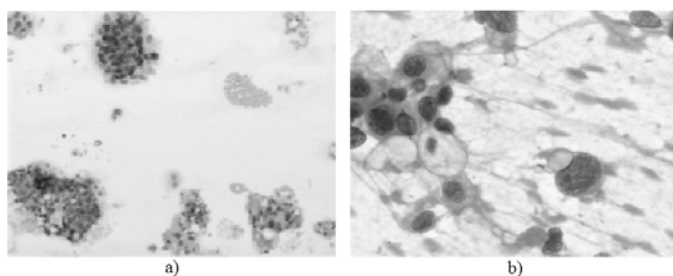


Fig. 6 FNA images: a) 100× resolution; b) 400× resolution.

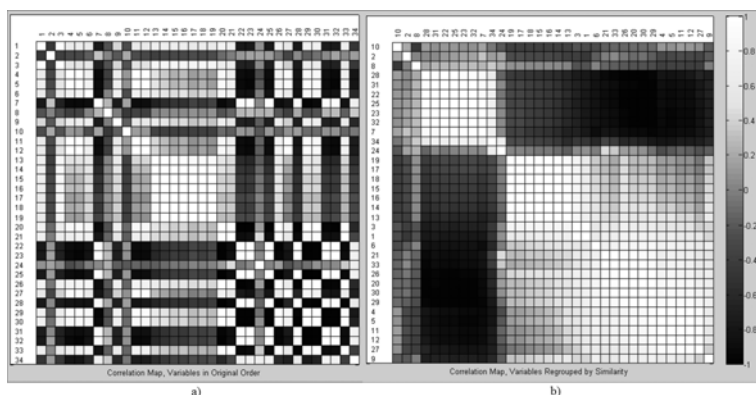


Fig. 7 Correlation between extracted features: a) with variables in original order; b) with variables regrouped by similarity.

average, for most tested classifiers, the best classifications were obtained when the level set segmentation was used.

The described work done by Jeleń *et al.* was applied to a classification system built and currently being tested in the pathological laboratory. In [100] the authors show that preliminary medical tests provide promising results and the automated breast cancer grading system performs with a high accuracy when applied to the real and unseen data. The achieved accuracy was 81.96%.

10 Challenges and Future Developments

There are many challenging problems in automatic processing of histopathologies. The main problems include large variation of blood and cancer cells, occlusions, segmentation, low quality of images and difficulties in getting real data. We believe that these difficulties will be overcome with time.

References

- [1] Newland, J.: The peripheral blood smear. In: Goldman, L., Ausiello, D. (eds.) *Cecil Medicine V*, ch. 161, 23rd edn., Saunders Elsevier, Philadelphia (2007)
- [2] Agby, G.: Leukopenia and leukocytosis. In: Goldman, L., Ausiello, D. (eds.) *Cecil Medicine*, ch. 173, 23rd edn., Saunders Elsevier, Philadelphia (2007)
- [3] Ramoser, H., Laurain, V., Bischof, H., Ecker, R.: Leukocyte segmentation and classification in blood-smear images. In: 27th IEEE Annual Conference Engineering in Medicine and Biology, Shanghai, China, September 1-4 (2005)
- [4] Al-Muhairy, J., Al-Assaf, Y.: Automatic white blood cell segmentation based on image processing. In: 16th IFAC World Congress (2005)
- [5] Prasad, B., Prasanna, S.M. (eds.): *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*. Studies in Computational Intelligence, vol. 83. Springer, Heidelberg (2008); ISBN 978-3-540-75397-1
- [6] Piuri, V., Scotti, F.: Morphological classification of blood leucocytes by microscope images. In: IEEE International Conference on Computational Intelligence Far Measurement Systems and Applications, Boston, MA, July 14-16 (2004)
- [7] Bentley, S., Lewis, S.: The use of an image analyzing computer for the quantification of red cell morphological characteristics. *British Journal of Hematology* 29, 81–88 (1975)
- [8] Rowan, R.: Automated examination of the peripheral blood smear. In: *Automation and quality assurance in hematology*, ch. 5, pp. 129–177. Blackwell Scientific, Oxford (1986)
- [9] Costin, H., Rotariu, C., Zbancioc, M., Costin, M., Hanganu, E.: Fuzzy rule-aided decision support for blood cell recognition. *Fuzzy Systems & Artificial Intelligence* 7(1-3), 61–70 (2001)
- [10] Albertini, M., Teodori, L., Piatti, E., Piacentini, M., Accorsi, A., Rocchi, M.: Automated analysis of morphometric parameters for accurate definition of erythrocyte cell shape. *Cytometry Part A* 52A(1), 12–18 (2003)
- [11] Robinson, R., Benjamin, L., Cosgri, J., Cox, C., Lapets, O., Rowley, P., Yatco, E., Wheelless, L.: Textural differences between aa and ss blood specimens as detected by image analysis. *Cytometry* 17(2), 167–172 (1994)
- [12] Gering, E., Atkinson, C.: A rapid method for counting nucleated erythrocytes on stained blood smears by digital image analysis. *Journal of Parasitology* 90(4), 879–881 (2004)
- [13] Martelli, A.: An application of heuristic search methods to edge and contour detection. *Communications of the ACM* 19(2), 73–83 (1976)
- [14] Fleagle, S., Johnson, M., Wilbricht, C., Skorton, D., Wilson, R., White, C., Marcus, M., Collins, S.: Automated analysis of coronary arterial morphology in cineangiograms: geometric and physiologic validation in humans. *IEEE Transactions on Medical Imaging* 8(4), 387–400 (1989)
- [15] Fleagle, S., Thedens, D., Ehrhardt, J., Scholz, T., Skorton, D.: Automated identification of left ventricular borders from spin-echo magnetic resonance images. *Investigative Radiology* 26(4), 295–303 (1991)
- [16] Gonzalez, R., Woods, R.: *Digital Image Processing (3rd Economy Edition)*. Prentice–Hall, Englewood Cliffs (2008)
- [17] Adjouadi, M., Fernandez, N.: An orientation-independent imaging technique for the classification of blood cells. *Particle & Particle Systems Characterization* 18(2), 91–98 (2001)

- [18] Di Ruberto, C., Dempster, A., Khan, S., Jarra, B.: Segmentation of blood images using morphological operators. In: 15th International Conference on Pattern Recognition (ICPR 2000), Barcelona, Spain, pp. 397–400. IEEE, Los Alamitos (2000)
- [19] Gauch, J.: Image segmentation and analysis via multiscale gradient watershed hierarchies. *IEEE Transactions on Image Processing* 8(1), 69–79 (1999)
- [20] Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 4, 321–331 (1988)
- [21] Ongun, G., Halici, U., Leblebicioglu, K., Atalay, V., Beksac, S., Beksac, M.: Automated contour detection in blood cell images by an efficient snake algorithm. *Nonlinear Analysis-Theory Methods & Applications* 47(9), 5839–5847 (2001)
- [22] Wang, X., He, L., Wee, W.G.: Deformable contour method: A constrained optimization approach. *International Journal of Computer Vision* 59(1), 87–108 (2004)
- [23] Ongun, G., Halici, U., Leblebicioglu, K., Atalay, V.V., Beksac, M., Beksac, S.: Feature extraction and classification of blood cells for an automated differential blood count system. In: *Proc. IJCNN*, vol. 4, pp. 2461–2466 (2001)
- [24] Leznray, O., Elmoataz, A., Cardot, H., Gougeon, G., Lecluse, M., Elie, H., Revenu, M.: Segmentation of cytological images using color and mathematical morphology. *European Congress of Stereology* 18(1), 1–14 (1999)
- [25] Ravi, B., Kumar, Danny, K., Joseph, Sreenivas, T.V.: Teager energy based blood cell segmentation. In: *International Conference on Digital Signal Processing*, vol. 2, pp. 619–622 (2002)
- [26] Ongun, G., Halici, U., Leblebicioglu, K., Atalay, V., Beksac, M., Beksac, S.: An automated differential blood count system. In: *IEEE Int. Conf. on Engineering in Medicine and Biology Society*, vol. 3, pp. 2583–2586 (2001)
- [27] Sinha, N., Ramakrishnan, A.: Automation of differential blood count. In: *Conf. on Convergent Technologies for Asia-Pacific Region*, vol. 2, pp. 547–551 (2003)
- [28] Comaniciu, D., Meer, P.: Cell image segmentation for diagnostic pathology. In: Suri, J., Setarehdan, S., Singh, S. (eds.) *Advanced Algorithmic Approaches to Medical Image Segmentation: state-of-the-art application in cardiology, neurology, mammography and pathology*, pp. 541–558. Springer, Heidelberg (2001)
- [29] Jiang, K., Liao, Q.-M., Dai, S.-Y.: A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering. In: *Int. Conf. on Machine Learning and Cybernetics*, vol. 5, pp. 2820–2825 (2003)
- [30] Hamghalam, M., Motameni, M., Kelishomi, A.E.: Leukocyte segmentation in giemsa-stained image of peripheral blood smears based on active contour. In: *International Conference on Signal Processing Systems*, pp. 103–106. IEEE Computer Society, Los Alamitos (2009)
- [31] Fodor, I., Kamath, C.: On denoising images using wavelet-based statistical techniques. Tech. rep., Lawrence Livermore National Laboratory, uCRL JC-142357 (2001)
- [32] Sendur, L., Selesnick, I.: Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Transactions on Signal Processing* 50(11), 2744–2756 (2002)
- [33] Sendur, L., Selesnick, I.: A bivariate shrinkage function for wavelet-based denoising. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, pp. 1261–1264 (2002)
- [34] Lim, J.S.: *Two-Dimensional Signal and Image Processing*. Prentice-Hall, Englewood Cliffs (1990)

- [35] Hong, V., Palus, H., Paulus, D.: Edge preserving filters on color images. In: Bubak, M., van Albada, G.D., Sloot, P.M.A., Dongarra, J. (eds.) ICCS 2004. LNCS, vol. 3039, pp. 34–40. Springer, Heidelberg (2004)
- [36] Babic, Z., Mandic, D.: An efficient noise removal and edge preserving convolution filter. In: 6th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service, vol. 2, pp. 538–541 (2003)
- [37] Nikolaou, N., Papamarkos, N.: Color reduction for complex document images. *International Journal of Imaging Systems and Technology* 19(1), 14–26 (2009)
- [38] Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Sixth International Conference on Computer Vision, January 4–7, pp. 839–846 (1998)
- [39] Papari, G., Petkov, N., Campisi, P.: Artistic edge and corner enhancing smoothing. *IEEE Transactions on Image Processing* 16(10), 2449–2462 (2007)
- [40] Ntogas, N., Veintzas, D.: A binarization algorithm for historical manuscripts. In: 12th WSEAS International Conference on Communications, pp. 41–51 (2008)
- [41] Sauvola, J., Pietikainen, M.: Adaptive document image binarization. *The Journal of the Pattern Recognition Society* 33(2), 225–236 (2000)
- [42] Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on System, Man and Cybernetics* 9(1), 62–66 (1979)
- [43] Habibzadeh, M., Krzyzak, A., Fevens, T., Sadr, A.: Counting of RBCs and WBCs in noisy normal blood smear microscopic images. In: *SPIE Medical Imaging* (February 2011)
- [44] Vincent, L.: Fast opening functions and morphological granulometries. *Image Algebra and Morphological Image Processing* 2300, 253–267 (1994)
- [45] Lin, Y.-C., Tsai, Y.-P., Hung, Y.-P., Shih, Z.-C.: Comparison between immersion-based and toboggan-based watershed image segmentation. *IEEE Transactions on Image Processing* 15(3), 632–640 (2006)
- [46] Duda, R., Hart, P., Stork, D.: *Pattern Classification*, 2nd edn. Wiley Interscience Publishers, Hoboken (2000)
- [47] Cheng, H., Shi, X., Min, R., Cai, X., Du, H.N.: Approaches for Automated Detection and Classification of Masses in Mammograms. *Pattern Recognition* 39(4), 646–668 (2006)
- [48] Bottema, M., Slavotinek, J.: Detection and Classification of Lobular and DCIS (small cell) Microcalcifications in Digital Mammograms. *Pattern Recognition Letters* 21(13–14), 1209–1214 (2000)
- [49] Cheng, H., Cui, M.: Mass Lesion Detection with a Fuzzy Neural Network. *Pattern Recognition* 37, 1189–1200 (2004)
- [50] Cheng, H., Wang, J., Shi, X.: Microcalcification Detection using Fuzzy Logic and Scale Space Approaches. *Pattern Recognition* 37(2), 363–375 (2004)
- [51] De Santo, M., Molinara, M., Tortorella, F., Vento, M.: Automatic Classification of Clustered Microcalcifications by a Multiple Expert System. *Pattern Recognition* 36(7), 1467–1477 (2003)
- [52] Grohman, W., Dhawan, A.: Fuzzy Convex Set-based Pattern Classification for Analysis of Mammographic Microcalcifications. *Pattern Recognition* 34(7), 1469–1482 (2001)
- [53] Zhang, P., Verma, B., Kumar, K.: Neural vs. Statistical Classifier in Conjunction with Genetic Algorithm Based Feature Selection. *Pattern Recognition Letters* 26(7), 909–919 (2005)

- [54] Wolberg, W., Mangasarian, O.: Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology. *Proceedings of National Academy of Science, USA* 87, 9193–9196 (1990)
- [55] Mangasarian, O., Setiono, R., Wolberg, W.: Pattern Recognition via Linear Programming: Theory and Application to Medical Diagnosis. In: *Large-Scale Num. Opt.*, pp. 22–31. SIAM, Philadelphia (1990)
- [56] Street, W.N., Wolberg, W.H., Mangasarian, O.L.: Nuclear Feature Extraction for Breast Tumor Diagnosis. In: *Imaging Science and Technology/Society of Photographic Instrumentation Engineers 1993 International Symposium on Electronic Imaging: Science and Technology*, San Jose, California, vol. 1905, pp. 861–870 (1993)
- [57] Street, N.: *Cancer Diagnosis and Prognosis via Linear-Programming-Based Machine Learning*. Ph.D. thesis, University of Wisconsin (1994)
- [58] Street, N.: Xcyt: A System for Remote Cytological Diagnosis and Prognosis of Breast Cancer. In: Jain, L. (ed.) *Soft Computing Techniques in Breast Cancer Prognosis and Diagnosis*, pp. 297–322. World Scientific Publishing, Singapore (2000)
- [59] Lee, K., Street, W.: A Fast and Robust Approach for Automated Segmentation of Breast Cancer Nuclei. In: *Proceedings of the Second IASTED International Conference on Computer Graphics and Imaging*, Palm Springs, CA, pp. 42–47 (1999)
- [60] Lee, K., Street, W.: Model-based Detection, Segmentation and Classification for Image Analysis using On-line Shape Learning. *Machine Vision and Applications* 13(4), 222–233 (2003)
- [61] Wolberg, W.H., Street, W.N., Mangasarian, O.L.: Breast Cytology Diagnosis Via Digital Image Analysis. *Analytical and Quantitative Cytology and Histology* 15, 396–404 (1993)
- [62] Wolberg, W.H., Street, W.N., Mangasarian, O.L.: Machine Learning Techniques to Diagnose Breast Cancer from Image-Processed Nuclear Features of Fine Needle Aspirates. *Cancer Letters* 77, 163–171 (1994)
- [63] Walker, H.J., Albertelli, L., Titkov, Y., Kaltsatis, P., Seburyano, G.: Evolution of Neural Networks for the Detection of Breast Cancer. In: *Proceedings of International Joint Symposia on Intelligence and Systems*, pp. 34–40 (1998)
- [64] Walker, H.J., Albertelli, L.: Breast Cancer Screening Using Evolved Neural Networks. In: *IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 1619–1624 (1998)
- [65] Nezafat, R., Tabesh, A., Akhavan, S., Lucas, C., Zia, M.: Feature Selection and Classification for Diagnosing Breast Cancer. In: *Proceedings of International Association of Science and Technology for Development International Conference*, pp. 310–313 (1998)
- [66] Estevez, J., Alayon, S., Moreno, L.: Cytological Breast Cancer Fine Needle Aspirate Images Analysis with a Genetic Fuzzy Finite State Machine. In: *Conference Board of the Mathematical Sciences, CBMS 2002*, pp. 21–26 (2002)
- [67] Bagui, S., Bagui, S., Pal, K., Pal, N.: Breast Cancer Detection using Rank Nearest Neighbor Classification Rules. *Pattern Recognition* 36(1), 25–34 (2003)
- [68] Weyn, B., van de Wouwer, G., van Daele, A., Scheunders, P., van Dyck, D., van Marck, E., Jakob, W.: Automated Breast Tumor Diagnosis and Grading Based on Wavelet Chromatin Texture Description. *Cytometry* 33, 32–40 (1998)
- [69] Schnorrenberg, F., Pattichis, C., Kyriacou, K., Schizas, C.: Detection of Cell Nuclei in Breast Cancer Biopsies using Receptive Fields. In: *IEEE Proceedings of Engineering in Medicine and Biology Society*, pp. 649–650 (1994)

- [70] Schnorrenberg, F., Pattichis, C., Kyriacou, K., Schizas, C.: Content-based Description of Breast Cancer Biopsy Slides. In: Proc. Intl. EuroPACS Mtg., pp. 136–140 (1996)
- [71] Schnorrenberg, F., Pattichis, C., Kyriacou, K., Vassiliou, M., Schizas, C.: Computer-aided Classification of Breast Cancer Nuclei. *Technology & Health Care* 4(2), 147–161 (1996)
- [72] Schnorrenberg, F., Tsapatsoulis, N., Pattichis, C., Schizas, C., Kollias, S., Vassiliou, M., Adamou, A., Kyriacou, K.: A modular neural network system for the analysis of nuclei in histopathological sections. *IEEE Engineering in Medicine and Biology Magazine* 19, 48–63 (2000)
- [73] Belhomme, P., Elmoataz, A., Herlin, P., Bloyet, D.: Generalized region growing operator with optimal scanning: application to segmentation of breast cancer images. *Journal of Microscopy* 186, 41–50 (1997)
- [74] Adams, R., Bischof, L.: Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, 641–647 (1994)
- [75] Beucher, S.: Segmentation d'images et morphologie mathématique. Ph.D. thesis, Ecole National Supérieur des Mines de Paris (1990)
- [76] Beucher, S., Meyer, F.: *Mathematical Morphology in Image Processing*, ch. 12, pp. 433–481. Marcel Dekker, New York (1992)
- [77] Lezoray, O., Elmoataz, A., Cardot, H., Gougeon, G., Lecluse, M., Revenu, M.: Segmentation of cytological images using color and mathematical morphology. In: *European Conference on Stereology*, Amsterdam, Netherlands, p. 52 (1998)
- [78] Schüpp, S., Elmoataz, A., Fadili, J., Herlin, P., Bloyet, D.: Image segmentation via multiple active contour models and fuzzy clustering with biomedical applications. In: *The 15th International Conference on Pattern Recognition, ICPR 2000*, Barcelona, Spain, vol. 1, pp. 622–625 (2000)
- [79] Bloom, H., Richardson, W.: Histological Grading and Prognosis in Breast Cancer. *British Journal of Cancer* 11, 359–377 (1957)
- [80] Mangasarian, O., Street, W., Wolberg, W.: Breast Cancer Diagnosis and Prognosis via Linear Programming. *Operations Research* 43(4), 570–576 (1994)
- [81] Cheng, H., Li, X., Riodan, D. S., Scrimger, J.N.: A Parallel Approach to Tubule Grading in Breast Cancer Lesions and its VLSI Implementation. In: *Fourth Annual IEEE Symposium on Computer-Based Medical Systems*, pp. 322–329 (1991)
- [82] Cheng, H., Wu, C., Hung, D.: VLSI for Moment Computation and its Application to Breast Cancer Detection. *Pattern Recognition* 31(8), 1391–1406 (1998)
- [83] MacAulay, M., Scrimger, J., Riodan, D., Cheng, H.: An Interactive Graphics Package with Standard Examples of the Bloom and Richardson Histological Grading Technique. In: *Fourth Annual IEEE Symposium on Computer-Based Medical Systems*, pp. 108–112 (1991)
- [84] Gurevich, I., Murashov, D.: Method for early diagnostics of lymphatic system tumors on the basis of the analysis of chromatin constitution in cell nucleus images. In: *The 17th International Conference on Pattern Recognition, ICPR 2004*, Cambridge, UK, pp. 806–809 (2004)
- [85] Florack, L., Kuijper, A.: The topological structure of scale-space images. *Journal of Mathematical Imaging and Vision* 12(1), 65–80 (2000)
- [86] Rodenacker, K.: Applications of topology for evaluating pictorial structures. In: *Theoretical Foundations of Computer Vision*, pp. 35–46. Akademie-Verlag, Berlin (1993)
- [87] Rodenacker, K.: Quantitative microscope image analysis for improved diagnosis and prognosis of tumors in pathology. In: *Creaso Info Medical Imaging*, Creaso GmbH, Gilching, vol. 22 (1995)

- [88] Rodenacker, K., Bengtsson, E.: A feature set for cytometry on digitized microscopic images. *Anal. Cell Pathol.* 25(1), 1–36 (2003)
- [89] Weyn, B., Van de Wouwer, G., Koprowski, M., et al.: Value of morphometry, texture analysis, densitometry and histometry in the differential diagnosis and prognosis of malignant mesothelioma. *Journal of Pathology* 4(189), 581–589 (1999)
- [90] Young, I., Verbeek, P., Mayall, B.: Characterization of chromatin distribution in cell nuclei. *Cytometry* 7(5), 467–474 (1986)
- [91] Gurevich, I., Kharazishvili, D., Murashov, D., Salvetti, O., Vorobjev, I.: Technology for automated morphologic analysis of cytological slides. methods and results. In: *The 18th International Conference on Pattern Recognition (ICPR 2006)*, Hong Kong, China, pp. 711–714 (2006)
- [92] Churakova, Z., Gurevich, I., Jernova, I., et al.: Selection of diagnostically valuable features for morphological analysis of blood cells. *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications* 13(2), 381–383 (2003)
- [93] Qineti, Q.: *Automated Histopathology Breast Cancer Analysis and Diagnosis System*. Data Sheet (2005), <http://www.qinetiq.com/>
- [94] Jeleń, Ł.: *Computerized Cancer Malignancy Grading of Fine Needle Aspirates*. Ph.D. thesis, Concordia University (2009)
- [95] Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M., Tomaszewski, J.: Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 284–287 (2008)
- [96] Jeleń, Ł., Fevens, T., Krzyżak, A.: Classification of breast cancer malignancy using cytological images of fine needle aspiration biopsies. *Int. J. Math. Comput. Sci.* 18, 75–83 (2008)
- [97] Jeleń, Ł., Krzyżak, A., Fevens, T.: Comparison of pleomorphic and structural features used for breast cancer malignancy classification. In: Bergler, S. (ed.) *Canadian AI. LNCS (LNAI)*, vol. 5032, pp. 138–149. Springer, Heidelberg (2008)
- [98] Jeleń, Ł., Fevens, T., Krzyżak, A., Jeleń, M.: Discriminatory power of cells grouping features for breast cancer malignancy classification. In: *Proceedings of the International Federation for Medical and Biological Engineering*, vol. 21(3), pp. 559–562. Springer, Heidelberg (2008)
- [99] Jeleń, Ł., Fevens, T., Krzyżak, A.: Influence of nuclei segmentation on breast cancer malignancy classification. In: *Proceedings of SPIE*, vol. 7260, pp. 726014–726014–9 (2009)
- [100] Jeleń, Ł., Lipiński, A., Detyna, J., Jeleń, M.: Clinical verification of computerized breast cancer malignancy grading. *Bio-Algorithms and Med-Systems* 6(12), suppl. 1, 81–82 (2010)

Part VI
MISCELLANEOUS APPLICATIONS

Two-Stage Data Reduction for a SVM Classifier in a Face Recognition Algorithm Based on the Active Shape Model

Maciej Krol and Andrzej Florek

Abstract. In this paper, two stage data reduction method for face identification with use of Support Vector Machine (SVM) classifier is evaluated. SVM Classification was performed for data acquired from contour description of 2200 faces of 100 persons. Face contours were extracted from frontal face images with use of Active Shape Model (ASM) method. Two stage PCA+LDA data reduction performance is measured in comparison with single stage PCA or LDA reductions. We propose to replace first stage PCA reduction with much simpler and less computationally intensive contour decimation.

1 Introduction

For over decade Active Shape Model (ASM) has been evaluated for face contour extraction [4],[5]. The extracted face contours can be used for task of person identification with use of statistical classifiers such as SVM [6]. We approach small sample size problem in SVM training with two stage data reduction PCA+LDA [1],[8] and compare it with single stage PCA or LDA reduction [3]. Replacing first stage PCA reduction by k -decimation method is investigated.

2 Face Contour

Face contours were extracted from frontal face images with use of ASM method. The shape in ASM is represented as an ordered set of N control points placed on

Maciej Krol

Institute of Control and Information Engineering, Poznan University of Technology, str. Piotrowo 3A, Box 1, 60-965 Poznan, Poland

e-mail: maciej.krol@interia.pl

Andrzej Florek

Institute of Control and Information Engineering, Poznan University of Technology, str. Piotrowo 3A, Box 1, 60-965 Poznan, Poland

e-mail: andrzej.florek@put.poznan.pl

face contours describing face elements (Fig. 1) and it is given by the following vector

$$\mathbf{x} = (x_1, x_2, \dots, x_N)^T \quad (1)$$

where x_i is a coordinate of shape control point expressed in common reference frame. Two subsequent vector elements x_{2j-1} and x_{2j} refer to horizontal and vertical coordinates of single control point ($1 \leq j \leq N/2$). Eight contours with total 194 points were arbitrary chosen (Tab. 1) for experiments. The feature vector \mathbf{x} in this case has $N = 388$ elements.

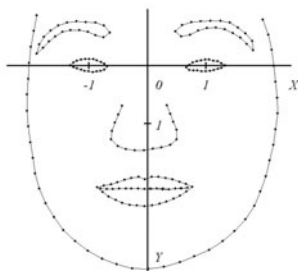


Fig. 1 Set of 8 contours extracted from face images

Table 1 Number of points in face contours

Face contour	Number of points
Face Outline FO	41
Mouth Outer MO	28
Mouth Inner MI	20
Right Eyelid REL	20
Left Eyelid LEL	20
Right Eyebrow REB	20
Left Eyebrow LEB	20
Nose Outline NO	17
TOTAL	194

2.1 Image Database

The PUT image database was used in contour fitting process [7]. The database contains set of images of 100 people. Shots were taken mainly from frontal view with limited head rotation and varying scale. Images contain faces with and without structural disturbances such as glasses, beard or long hair. Each person has at least 100 pictures in different head orientation. Three sequences of pictures based on the head orientation and face expression were selected for classification:

- H - sequence of frames for horizontal head rotation from right to left profile,
- V - sequence of frames for vertical head rotation from facing up to down,
- D - frames for varying head position and facial expressions.

Manually placed contours were used as a learning set (LSM) for ASM. Contours were manually placed on images of 100 people. For each person 11 images from H sequence and 11 images from V sequence were described with contours. Set of 2200 images with manually placed contour was used for training the ASM.

2.2 Active Shape Model

The ASM uses two statistical distributions to model contours on the image.

- Point Distribution Model - models statistical distribution of position of control points.
- Local Grey Level Model - models statistical distribution of image values along normal direction to contour in vicinity of control points.

The PDM uses Principal Component Analysis (PCA) approximation of shape of contours. The P matrix contains t eigenvectors corresponding to the largest values of the covariance matrix C_x . The covariance C_x and mean \bar{x} was computed from 2200 manually placed contours x_i .

The shape vector x_i of the i -th sample (contour on image) can be approximated by the reduced shape vector b_i

$$b_i = P^T(x_i - \bar{x}) \quad (2)$$

being the result of the PCA method applied to covariance matrix C_x .

It is possible to approximate x_i with the given reduced shape vector b_i ,

$$x_i \approx \bar{x} + P b_i \quad (3)$$

2.3 Contour Fitting

Face contours are fitted to arbitrary face image with use of ASM algorithm. The implementation diagram is presented on figure 2. First step of algorithm is the face and eye localization based on template matching. Next, a shape model is initialized by placing the mean face shape model at the face image. Mean shape expressed in local coordinate system is translated, rotated and scaled to image coordinate system in such a way that centers of eye contours are positioned at eye pupils with on the image. Eye pupils are located in local coordinate system at $(-1, 0)$ and $(1, 0)$ respectively for right and left eye (Fig. 1). Final contours are uniformly resampled because control points have been shifted into neighbourhood of the average shape \bar{x} during ASM fitting procedure (LGL localization on Fig. 2).

The ASM algorithm consists of two stages. The first stage, the shape model is deformed to fit the image function. The LGL model is used to determine the best fit

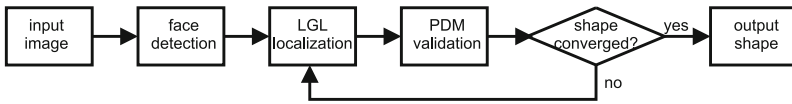


Fig. 2 Contour fitting algorithm

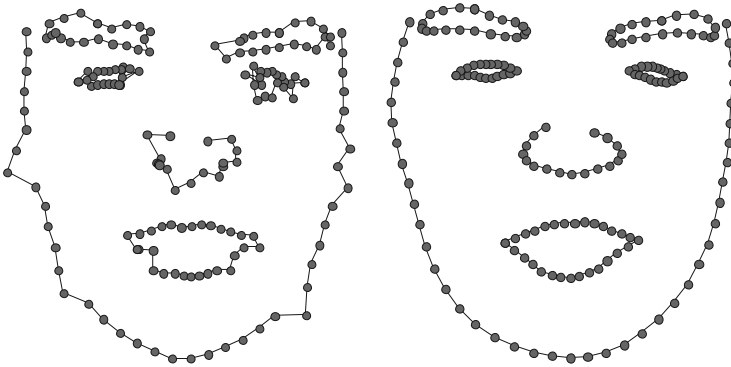


Fig. 3 Contour before (left) and after (right) PDM validation stage



Fig. 4 Fitted contours for different size of reduction space $t=15$ (left); $t=30$ (middle); $t=55$ (right)

function value. The second stage of the ASM algorithm is a PDM validation. At this stage, forward (2) and reverse (3) space transformation are consecutively applied. This is to reduce the distortions of fitted shape from the model (Fig. 3). It gives an effect of attenuation of the small noise in local shape position. Two above mentioned stages are executed one after another until the changes of position of points between previous and current step of algorithm are small enough. The result of the algorithm is a shape model fitted to a face on the image (Fig. 4).

The accuracy of contour fitting depends on size of reduction space t (Fig. 4). Small value of t makes contours rigid in respect to mean shape \bar{x} . Differences between people are dampened. Large value of t causes relaxation of contours but makes them susceptible to structural disturbances or noise. The problem of selection of t will be further investigated in section 3.4.

3 Classification

Contour fitted with ASM algorithm are used as learning and testing sets in a statistical classification. We evaluated multi-class SVM for person identification within database of 100 people.

3.1 Contour Set Definitions

For training and testing the classifiers two data sets were defined:

- LSN - normalized automatic learning set, 2200 contours, 100 classes, 22 contours for each class (11 central from H-sequence and 11 central from V-sequence);
- TBS - testing boundary set, 1100 normalized automatic contours, 100 classes, 11 contours for each person: four from H-sequence, four from V-sequence (in boundary case face position - just beyond position chosen to LSN) and three from D-frame set.



Fig. 5 Contours from LSN (top row) and TBS (bottom row) fitted by ASM

3.2 SVM Training

In the presented approach SVM classifiers with one-to-one voting system were tested [2]. Two types of SVM kernel functions $K(\mathbf{x}_i, \mathbf{x}_j)$ were evaluated:

- Linear Function: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
- Radial Basis Function (RBF): $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2); \gamma > 0$

Classification was performed for contour representation (II) in the full feature space (size of 388) and in its p -subspaces obtained after the PCA or LDA reduction. The value of parameter C (the penalty parameter of a soft margin of SVM classifier)

and γ in case of RBF were selected by grid search on learning set LSN. Pairs of (C, γ) were tested and the one with best cross-validation accuracy was used to narrow down the search on the grid of C and γ . In cross-validation procedure LSN was subdivided into 5 subsets of equal amount of samples for each class (person identity). The SVM was trained sequentially on 4 subsets and tested with one remaining subset. Cross validation accuracy was measured by classification rate which is defined as ratio of number of correctly identified samples to total number of samples. The optimization procedure was performed in the full feature space three times for three different data scaling strategies - no scaling, scaling in the range of $[0, 1]$ and scaling in the range of $[-1, 1]$.

After extensive grid search parameters $C = 32$, $\gamma = 0.016$ and linear scaling of each attribute to the range of $[-1, 1]$ were picked as suboptimal ones. We found out that kernel type (linear or RBF), optimal scaling factors and SVM parameters C and γ do not change substantially when optimization is performed for different SVM kernel or in reduced feature spaces (sec. 3.4). Thus, for the sake of clarity, tests presented in this paper were performed with above mentioned configuration parameters.

3.3 SVM Testing

Verification of the SVM classifier on the set which was trained (LSN) gives overall good results (classification rate ca. 95%). We stressed the classifier with set of boundary cases (TBS). The classification rate in case of TBS plummeted down to about 60%. In our case scaling data to range $[-1, 1]$ yielded better classification results than scaling to $[0, 1]$ or without scaling (Fig. 6).

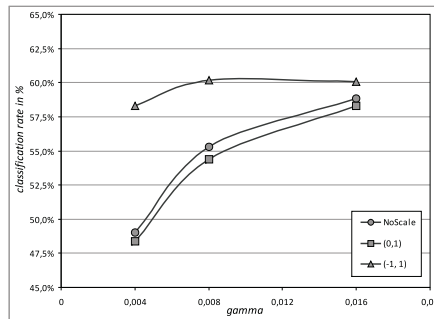


Fig. 6 Influence data scaling in classification rate of SVM-RBF ($C=32$) in TBS set

Low classification rate for TBS set led us to analyze the impact of various parameters on the final result. Influence of size of t used in contour validation in contour fitting stage, types and size of transformation from contour space to the classification feature space are investigated further in the paper.

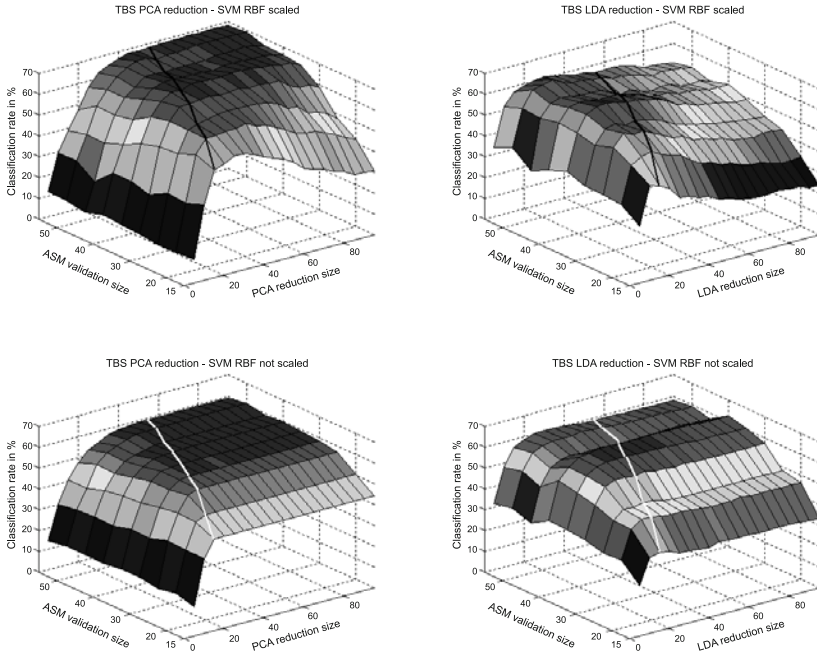


Fig. 7 SVM classification rate for different reduction methods. Line $m = t$ is marked on the plots.

3.4 Feature Space Reduction

We tested two common approaches to feature space reduction to size m -transformation to PCA and LDA subspaces. In PCA eigenvalue decomposition of covariance matrix of LSN is calculated. In case of LDA generalized eigenvalue decomposition from within-class and between-class covariance matrices is performed. The problem of singularity of covariance matrices is particularly salient for within-class covariance matrix used in LDA method. The LSN set contains only 22 contours for each classes in a set of 100 classes. Additionally the feature vector \mathbf{x}_i of size 388 generated by ASM contour fitting has non zero variance only in t directions due to PDM validation step which is simply forward and inverse PCA transformation constrained to t directions. Hence, the rank of within-class covariance is not larger than t .

We examined influence of t and m on SVM classification rate by performing test for selected permutations of (m, t) pairs. The size of t is constrained by singularity of PCA decomposition in ASM contour fitting. The size of m is constrained by singularity of PCA or LDA decomposition in feature space reduction step. The classification rate of SVM with RBF kernel for not scaled and scaled data in range $[-1, 1]$ are presented on figure [7](#).

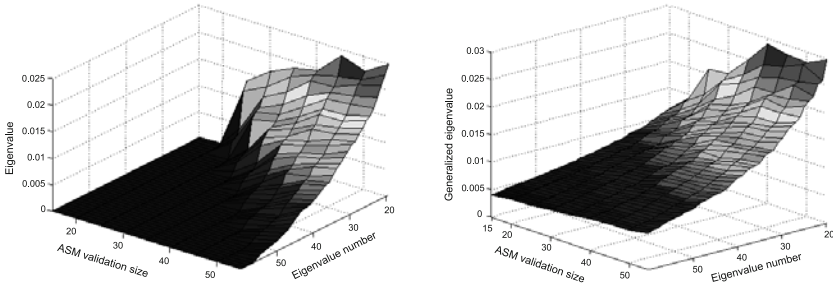


Fig. 8 Eigenvalues of PCA (left) and generalized eigenvalues of LDA (right).

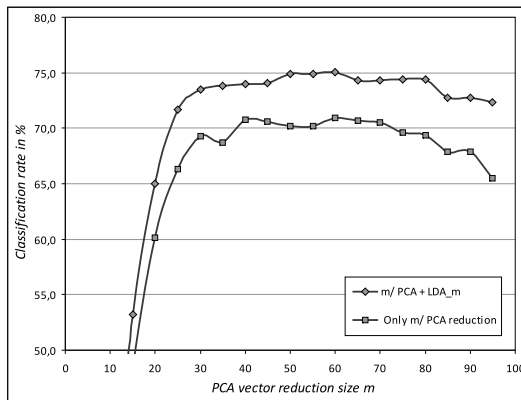


Fig. 9 Comparison of classification rate between single stage PCA and two stage PCA+LDA data reduction to size m for scaled data and SVM-RBF $t=35$ classifier

Reduction in PCA outperforms LDA reduction. Data scaling in case of single stage reduction size exceeding PDM validation size ($m > t$) decreases classification rate. It is probably caused by rounding errors and amplification of noise generated in normalization and resampling procedures (Fig. 8). For $t > 35$ the increase of classification rate is negligible.

Two stage method of feature space reduction was examined. In the first stage PCA reduction to size m is performed in the next stage LDA reduction to size p is applied. This method is known as PCA+LDA [8], [9]. The result of classification for $t = 35$ and $p = m$ are presented on figure 9. Additional LDA reduction after PCA yields increase of classification rate by 5% in comparison to single stage PCA reduction. We observed that two stage reduction with $m > t$ improves classification rate. Thus for SVM classifier using directions of feature vector \mathbf{x} with PDM eigenvalues equal or close to zero is desired in small quantities. It can be interpreted as adding to feature vector components of equal value for every class (positioned on the average

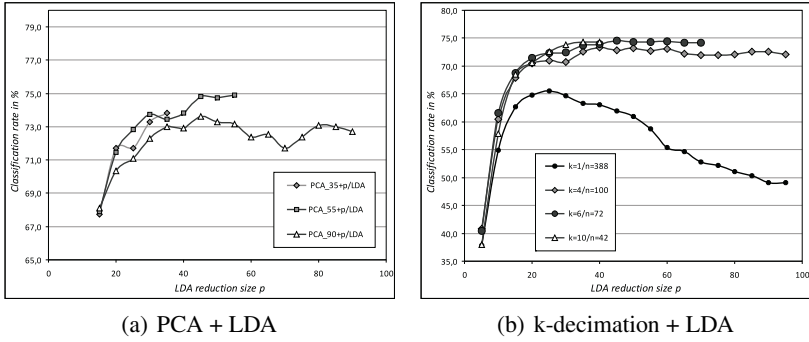


Fig. 10 Classification rate in function of LDA reduction p for SVM-RBF classifier $t=35$

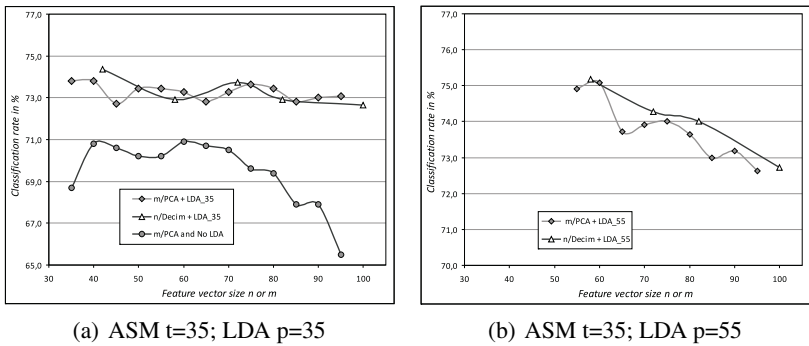


Fig. 11 Comparison between PCA, PCA+LDA and k -decimation+LDA reduction efficiency for scaled data and SVM-RBF classifier (k -decimation reduces vector to size n)

contour \bar{x}). An improvement of classification rate with increase of parameter m is not unbounded and has a certain optimal value (Fig. 9 and 10(a)).

Good classification rates were obtained by replacing first stage PCA reduction with much simpler k -decimation. In k -decimation every k -th control point of contour was selected to reduced vector (Fig. 10(b) and 11). Contours constituting feature vector (Tab. 1) were decimated separately hence size of reduced feature space n is not always equal to N/k .

4 Summary

Several feature space reduction method for SVM classification of contours obtained from ASM were evaluated. Interaction between PDM validation in ASM contour fitting and feature space reduction in SVM classification was analyzed. In reduction of feature vector x to size m we propose to set $m > t$ that is above the number of features t in reduced contour vector b of PDM validation. Two stage PCA+LDA

reduction prior classification performs better than single stage PCA or LDA reduction. First stage PCA reduction can be successfully replaced by simpler and less computationally intensive k -decimation method (Fig. 10 and 11).

References

- [1] Belhumeur, P., Hespanha, J., Kriegman, D.J.: Eigenfaces versus fisherface: recognition using class specific linear projection. *IEEE Trans. Pattern Analysis Machine Intelligence* 19(7), 711–720 (1997)
- [2] Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines (2001), <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [3] Chen, L., Liao, H., Ko, M., Lin, J., Yu, G.: A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition* 33(10), 1713–1726 (2000)
- [4] Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models - their training and application. *Computer Vision and Image Understanding* (61), 38–59 (1995)
- [5] Cootes, T.F., Taylor, C.J.: *Statistical Models of Appearance for Computer Vision*. Technical report, Imaging Science and Biomedical Engineering, University of Manchester (2001)
- [6] Vapnik, V., Cortes, C.: *Support-Vector Networks*. *Machine Learning* (20) (1995)
- [7] Kasinski, A., Florek, A., Schmidt, A.: The PUT face database. *Image Processing and Communications* 13(3-4), 59–64 (2008)
- [8] Yu, H., Yang, J.: A direct LDA algorithm for high-dimensional data – with application to face recognition. *Pattern Recognition* 34(10), 2067–2070 (2001)
- [9] Zhang, S., Sim, T.: Discriminant Subspace Analysis: A Fukunaga-Koontz Approach. *IEEE Trans. Pattern Analysis Machine Intelligence* 29(10), 1732–1745 (2007)

A New Star Identification Algorithm Based on Fuzzy Algorithms

Shahin Sohrabi and Ali Asghar Beheshti Shirazi

Abstract. The proposed algorithm is for satellite attitude determination. In this algorithm, the star point pattern convert to line pattern by "Delaunay triangulation" method, then we present a fuzzy line pattern matching. In this method, we use the membership functions to describe position, orientation and relation similarities between different line segments. The simulation results based on the "Desktop Universes Star images" demonstrate that the fuzzy star pattern recognition algorithm speeds up the process of star identification and increases the rate of success greatly (96.4%) compared with traditional matching algorithms. In addition, since the quality of star images play an important role in improving accuracy of star pattern recognition algorithm, therefore for image pre-processing we propose a fuzzy edge detection technique. This method highly affects noise cancellation, star features extraction, database production and matching algorithm.

1 Introduction

Star observation is widely used by spacecraft for attitude determination. Star trackers determine the attitude of the spacecraft from some reference stars in the celestial sphere. The process of determination attitude can be divided into four steps. First, the star tracker takes the star images using the image detectors (CCD or CMOS) in an arbitrary direction. Secondly, the microprocessor operates the image processing algorithm on the star images and enhances the quality of them. Thirdly, the star features are extracted based on the own matching strategy and stored in the star database and then the star constellations are matched to a star database using by the

Shahin Sohrabi

Iran University of Science and Technology

e-mail: sohrabishahin@vu.iust.ac.ir

Ali Asghar Beheshti Shirazi

Engineering group, Iran University of Science and Technology

e-mail: abeheshti@iust.ac.ir

star pattern recognition (SPR) algorithm. Finally, the attitude matrix and attitude angle (Roll, Pitch, Yaw) are calculated to estimate to the attitude control system.

The third part of the process is a most crucial step, because the accuracy identification and time consumed are the most important parameters in performance of the star identification algorithm. There are different star pattern identification algorithms according to different strategies for autonomous attitude determination, such as, triangle matching algorithm [1], grid matching algorithm [2] and [3]. Recently, other algorithms, for examples, singular value decomposition (SVD) algorithm [4] are also applied in this field. Some algorithms apply new neural network techniques [5] and genetic algorithm [6] and fuzzy decision method [7], but the algorithms are too complex to be used in the star sensors on-board a spacecraft. Therefore to reduce the computing workload of the star identification, speed up the identification process and guarantee high rate of success, a new star identification algorithm based on fuzzy line pattern matching is proposed in this paper.

In this work, we focus on the third parts of the process of determination attitude that described above. Therefore, after image processing operations and star database generation, the stars point pattern is converted to a line pattern by "De-launay triangulation" method. Then we consider three fuzzy sets, line location, line orientation and directional relationship between lines and make use of the corresponding three membership functions for line similarity measurement. Finally, for star pattern recognition we define a cost matching function for matching the star line patterns to star database patterns. In addition, the quality of star images highly affects star features extraction, database production and matching algorithm. So before star pattern recognition (SPR) algorithm, the star images is processed by image processing algorithm that can be used for improving accuracy of star pattern recognition include noise cancellation and star features extraction. Therefore, we use the optimum Gaussian filter based on fuzzy edge detection for noise cancelling. There are many edge detection methods such as Sobel, Canny and Robert [8], [9], [10]. But the simulation results show that the fuzzy edge detection only modifies noisy pixels and has the minimum destructive effect on the edges of stars.

This paper is separated into five major sections: the pre-processing algorithm that consists of noise cancellation and blurs removal is presented in second section, and then star identification algorithm based on fuzzy line pattern matching is introduced in the third section. In last two sections, the simulations are presented and the results are discussed.

2 Image Pre-processing

The quality of star images highly affects star features extraction, database production and matching algorithm. Therefore noise cancellation will play an important role in improving accuracy of star pattern recognition algorithm.

Since the first step in features extraction is edge detection, the noises cancellation must have the minimum destructive effect on the edges of stars. In this work, the optimum Gaussian filter is used for noise cancellation. This filter only modifies

noisy pixels. To select the noisy pixels, as the first step, we calculate the gradient of the image by using fuzzy edge detection method. Then we calculate the centroid of the histogram of the edges. We consider all pixels as noisy pixels, which has the gradient value less than the centroid value and then apply the noise cancellation algorithm. To calculate the centroid, assume that the gray levels values is g_0, g_1, \dots, g_n respectively. Where $g_0 < g_1 < \dots < g_n$, and for any one gray value ($i = 1, 2, \dots, n$) the number of pixels with their gray values equal to g_i is q_i , Then the centroid is obtained as following

$$Q = \begin{cases} \frac{\sum_{i=1}^n g_i q_i}{\sum_{i=1}^n q_i} & g_0 = 0 \\ \frac{\sum_{i=0}^n g_i q_i}{\sum_{i=0}^n q_i} & g_0 = 0 \end{cases} \tag{1}$$

By using the Q threshold the star image is divided in to two black and white classes (A_b, A_w). The background of image will be a membership of one of them. Now, the centroid of the black and white area is obtained by using the histogram graph.

The Purpose of calculation of A_w, A_b is finding a standard way that is used for reduction the dispersal of the histogram graph. Thus, after determination of the image background, the histogram graph converges in to A_b or A_w . the process of this method will be explained. A_w, A_b Is calculated by following mapping function G , then the image is transferred from spatial domain to fuzzy domain.

$$p_{i,j} = G(f_{i,j}) = \begin{cases} \frac{f_{i,j} - f_{\min}}{A_w - f_{\min}} & f_{i,j} \leq A_b \\ \frac{A_w - f_{i,j}}{A_w - A_b} & A_b \leq f_{i,j} \leq A_w \\ \frac{f_{\max} - f_{i,j}}{f_{\max} - A_w} & f_{i,j} \geq A_w \end{cases} \tag{2}$$

$$i = 1, 2, \dots, M; j = 1, 2, \dots, N.$$

Where f_{\min} and f_{\max} denote the maximal gray value and the minimal one in the image. M and N denote the rows and columns of the image, respectively. When the image is changed from the spatial domain into the fuzzy domain, the fuzzy enhancement H_r operator is adopted to realize the function of image enhancement as follows

$$p'_{i,j} = H_r^{(p_{i,j})} = H_1(H_{r-1}(p_{i,j})) \tag{3}$$

$$H_r^{(p_{i,j})} = \begin{cases} \frac{p_{i,j}^2}{t} & 0 \leq p_{i,j} \leq t \\ \frac{1 - (1 - p_{i,j})^2}{1 - t} & t \leq p_{i,j} \leq 1 \end{cases} \tag{4}$$

Where r denotes the iterative times, and it has a great influence on the image enhancement effect. To enhance the image moderately, is usually chosen as 2 or 3. t denotes the fuzzy characteristic threshold, and it can be chosen more flexibly in this algorithm than the traditional fuzzy enhancement algorithm. To select the value of t , the distribution of the histogram graph and the class of background of image must be considered. Indeed, when t is close to 0, the histogram graph will go from

A_b to A_w and vice versa. In this paper, we assume $r = 3, t = 0.5$. Thereafter, the inverse mapping transformation G^{-1} transfers the image from fuzzy domain to spatial domain as follows

$$h_{i,j} = G^{-1}(p'_{i,j}) = \begin{cases} (A_b - f_{\min})p'_{i,j} + f_{\min} & f_{i,j} \leq A_b \\ A_w - (A_w - A_b)p'_{i,j} & A_b \leq f_{i,j} \leq A_w \\ (f_{\max} - A_w)p'_{i,j} + A_w & f_{i,j} \geq A_w \end{cases} \quad (5)$$

$$i = 1, 2, \dots, M; j = 1, 2, \dots, N$$

A_{bck}, A_{obj} Indicates the centroid of the background (night sky) and the objects (stars) respectively. Ultimately, after changing the color of the pixels (see Fig.1), we see that the image is divided in to two classes with the same color A_{bck}, A_{obj} . Also, because of existing the same colors in the enhanced image, the edges will be more continual.

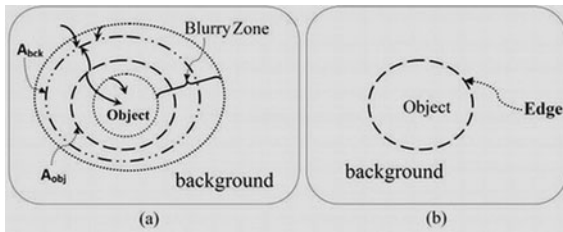


Fig. 1 The image enhancement algorithm.(a) The propose algorithm (b) The result

3 Fuzzy Star Pattern Recognition

Many pattern matching problems can be simplified as a point or line pattern matching task [11]. In this work, we consider fuzzy algorithm for matching line patterns. A point pattern can be converted to a line pattern by connecting the points with line segments, so line matching algorithms can be applied to point matching after a proper point to line transformation. Therefore, as first step, the stars point pattern is converted to a line pattern by "Delaunay triangulation" method. This section will initially specify a description of similarity measures between line segments and matching algorithm. The process of database generation will be given in detail. The matching problem we need to solve here can be stated as follows. We are given a set of star pattern (prototype pattern).

$$P = \{p^{(1)}, p^{(2)}, \dots, p^{(N_p)}\} \quad (6)$$

Where N_p is the number of star patterns and each star pattern is $p^{(k)}$ consists of M_k line segments

$$p^{(k)} = \{l_1^{(k)}, l_2^{(k)}, \dots, l_{M_k}^{(k)}\} = \{(s_1^{(k)}, e_1^{(k)}), (s_2^{(k)}, e_2^{(k)}), \dots, (s_{M_k}^{(k)}, e_{M_k}^{(k)})\} \quad (7)$$

$$l_j^{(k)} = (s_j^{(k)}, e_j^{(k)})$$

Represents the starting and ending locations respectively line j in star pattern k . A database pattern q consisting of N line segments is denoted as

$$q = \{l_1, l_2, \dots, l_{M_k}\} = \{(s_1, e_1), (s_2, e_2), \dots, (s_N, e_N)\}, l_j = \{s_j, e_j\} \quad (8)$$

Represents the starting and ending locations respectively of line j in the database pattern.

3.1 Similarity Measures between Line Segments

The key problem in matching two line patterns is that for each line in one pattern we need to find the closest line in the other pattern. To carry out this task, we need to define how to measure the similarity between a pair of lines in two patterns. We consider two fuzzy sets, line location and line orientation between lines and make use of the corresponding two membership functions for line similarity measurement.

3.1.1 Location Similarity

We can treat the mid-point of a line as its location. The distance between the locations of line i in star pattern k and line j in the database pattern is shown in Fig.2. and can be calculated as

$$d_{ij}^{(k)} = \left| \frac{s_i^{(k)} + e_i^{(k)}}{2D^{(k)}} - \frac{s_j + e_j}{2D} \right| \quad (9)$$

Where $D^{(k)}$ and D are two normalization factors. In our experiments, we chose them to be the maximum width or maximum height, whichever is larger, of the star pattern and the database pattern respectively.

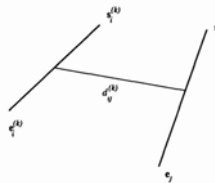


Fig. 2 Distance between lines i in star pattern k and line j in the database pattern defined as the distance between the mid-points of the two lines.

Assuming

$$s_i^{(k)} = \begin{bmatrix} x_{si}^{(k)} \\ y_{si}^{(k)} \end{bmatrix}, e_i^{(k)} = \begin{bmatrix} x_{ei}^{(k)} \\ y_{ei}^{(k)} \end{bmatrix}, s_j = \begin{bmatrix} x_{sj} \\ y_{sj} \end{bmatrix}, e_j = \begin{bmatrix} x_{ej} \\ y_{ej} \end{bmatrix} \quad (10)$$

We have

$$d_{ij}^{(k)} = \frac{1}{2} \sqrt{\left(\frac{y_{si}^{(k)} + y_{ei}^{(k)}}{D^{(k)}} - \frac{y_{sj} + y_{ej}}{D}\right)^2 + \left(\frac{x_{si}^{(k)} + x_{ei}^{(k)}}{D^{(k)}} - \frac{x_{sj} + x_{ej}}{D}\right)^2} \quad (11)$$

We can adjust the coordinate system so that

$$\min\{x_{si}^{(k)}, x_{ei}^{(k)}\} = \min\{y_{si}^{(k)}, y_{ei}^{(k)}\} = \min\{x_{sj}, x_{ej}\} = \min\{y_{sj}, y_{ej}\} = 0$$

$$D^{(k)} = \max\{x_{si}^{(k)}, x_{ei}^{(k)}, y_{si}^{(k)}, y_{ei}^{(k)}\}, D = \max\{x_{sj}, x_{ej}, y_{sj}, y_{ej}\} \quad (12)$$

The location similarity between the two lines can be measured using the following membership function

$$\mu_{L_{ij}}^{(k)} = \cos^{n_L} (c_L d_{ij}^{(k)}) \quad (13)$$

Where n_L and c_L are constants. In our experiments, we chose $n_L = 3$ and $n_L = \frac{\pi}{2}$. The minimum value of $d_{ij}^{(k)}$ is 0 and the maximum value is 1, corresponding to membership values of 1 and 0 respectively. In general, two lines have a large membership value if they are close to each other and vice versa.

3.1.2 Orientation Similarity

The difference between the orientations of a pair of lines plays an important role in matching the two lines. We define the membership function to measure the orientation similarity between line i in star pattern k and line j in the database pattern in terms of the angle, $\theta_{ij}^{(k)}$ between the two lines as follows

$$\mu_{O_{ij}}^{(k)} = \cos^{n_O} (c_O \theta_{ij}^{(k)}) \quad (14)$$

Where n_O and c_O are two constants, and

$$\theta_{ij}^{(k)} = \arg(s_i^{(k)} - e_i^{(k)}) - \arg(s_j - e_j) \quad (15)$$

Where \arg means the argument or phase of a vector. In our experiments, we chose $n_O = 3$, and $c_O = 1$. The angle is measured anti-clockwise from line i in star pattern k to line j in the database pattern. The membership value does not change if the

angle is measured anti-clockwise from line j in the database pattern to line i in star pattern k . In general,

$$0 \leq \theta_{ij}^{(k)} \leq \pi \tag{16}$$

So the membership value is in the range of $[0, 1]$. Two lines have a large membership value if they have similar orientations and vice versa.

3.2 Matching Algorithm

Our task here is to compare each star pattern with the database pattern so that the correspondence relations between the lines in a star pattern and the database pattern are identified and the best matching star pattern is determined. We make use of three similarity measures described in the preceding section to match a pair of lines. This can be done using the Shannon’s entropy function. Since we need the matching cost to decrease as the similarity measure increases, we make use of only the second half of the entropy function, which is a monotonic decreasing function for $0.5 < \mu < 1$.

$$H(\mu) = -(0.5 + 0.5\mu)\ln(0.5 + 0.5\mu) - (0.5 - 0.5\mu)\ln(0.5 - 0.5\mu) \tag{17}$$

Now $H(\mu)$ is a monotonic decreasing function for $0 < \mu < 1$, which can now be used as a cost function if we treat μ as the membership value of a similarity measure.

The cost for matching line i in star pattern k and line j in the database pattern can be defined as the weighted summation of the two components corresponding to two similarity measures

$$C_{ij}^{(k)} = a_L C_{Lij}^{(k)} + a_O C_{Oij}^{(k)} \tag{18}$$

Where $a_L = a_O = a_R = 1$ in our experiment. Finally, the cost for matching star pattern k and the database pattern is

$$C^k = \frac{1}{M_k} \sum_{i=1}^{M_k} \min_{j=1}^N C_{ij}^{(k)} \tag{19}$$

The line in database pattern that best matches line i in the star pattern is

$$j_i = \arg \min_{j=1} C_{ij}^{(k)} \tag{20}$$

4 Star Database Generation

In order to find stars rapidly, the star features are extracted and stored in the star database.

After image pre-processing, we converted the star point pattern to line pattern by "Delaunay triangulation" method. This process is shown in Fig.3. Then to star

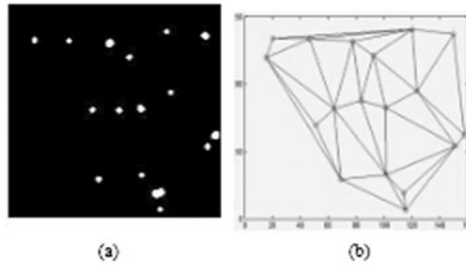


Fig. 3 (a)The guide stars. (b)The Delaunay triangulation

database generation the similarity measures are calculated as description in the preceding section and stored in the star database.

5 Simulation Results

We performed experiments in MATLAB environment, the simulation results based on "Desktop Universe Star Images" [12] is done in two steps. First, the image processing and guide stars selection is performed (see Fig.4). Then we apply the proposed algorithm on the star line pattern to compare with the star database pattern.

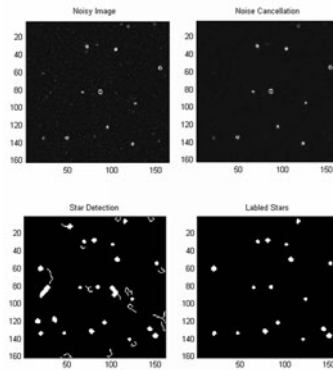


Fig. 4 The results of the proposed star image processing algorithm. (a) The noisy star image. (b) Noise cancellation. (c) Star detection. (d) Labeled Stars.

Table 1 Simulation results between algorithm in this paper and triangle algorithm

Algorithm	Recognition reliability(%)	Recognition time(ms)
Triangle Algorithm	91.30	1525
The Proposed Algorithm	96.40	1756

For the convenience, the performance of this method and the triangular pattern algorithm is compared in Tab.1. This simulation is repeated for 50 star images (128×128 pixel).

6 Conclusions

Efficient line matching procedures are useful for solving many pattern recognition problems. In this paper, we have presented a fuzzy line matching algorithm for star pattern recognition. Fuzzy algorithms can be used effectively for line pattern matching since they can deal with many ambiguous or fuzzy features of noisy star line patterns. In this method, we use the membership functions to describe position, orientation similarities between different line segments which are obtained from the star point patterns. Combining these similarities, we can match a line in a star pattern with a line in the database patterns if the two lines have similar measurements in the patterns. Experiment result demonstrates the efficiency and accuracy of the proposed algorithm is higher, and the memory requirement is lower, compared with the traditional star pattern recognition algorithm.

References

- [1] Scholl, M.S.: Star-field identification for autonomous attitude determination. *Star. J. Guidance, Control and Dynamics* 18, 61–65 (1995)
- [2] Padgett, C., Deldado, K.K.: A Gird Algorithm for Autonomous star identification. *IEEE Transactions on Aerospace and Electronic System* 33(1) (1997)
- [3] Meng, N., Danian, Z., Peifa, J., Tsinghua Univ, B.: Modified Grid Algorithm for Noisy All-Sky Autonomous Star Identification. *IEEE Transactions on Aerospace and Electronic Systems* 45(2), 516–522 (2009)
- [4] Jang, J.N., Kim, M.Y.: An Efficient and Robust Singular Value Method for Star recognition and attitude determination. *Journal of Astronautical Science* 52, 211–220 (2004)
- [5] Li, C.Y., Li, K., Zhang, Y.L.: Star pattern recognition based on neural Network. *Chinese Science Bulletin* 48(9), 892–895 (2003)
- [6] Paladugu, L., B.G.W., Schoen, M. P.: Star pattern Recognition for attitude determination using genetic algorithms. In: 17th AIAA-USU Conference on Small Satellites Logan (2003)
- [7] Zhang, H.: Fuzzy Decision-Based Star Identification Algorithm. In: *International Symposium on Computational Intelligence and Design*, vol. 1, pp. 116–119 (2008)
- [8] Pratt, W.: *Digital Image Processing*, 4th edn. Wiley-Interscience, NJ (2007)
- [9] Wang, Q., Ruan, H.B.: Fast fuzzy edge-detection algorithm. *J. Image Graph.* 6(1), 92–95 (2001)
- [10] Zhou, D.L., Pan, Q.: An improved algorithm of edge detection based on fuzzy sets. *J. Image Graph* 6(4), 353–358 (2001)
- [11] Van Wamelen, P.B., Iyengar, S.S.: A fast expected time algorithm for the 2-D point pattern matching problem. *The Journal of the Pattern Recognition Society* 37, 1699–1711 (2004)
- [12] Starry Night store, “DVDs”, <http://store.starrynight.com>

User Authentication Based on Keystroke Dynamics Analysis

Agata Kolakowska

Abstract. In this paper a few methods used to authenticate users on the basis of their keystroke dynamics have been presented. Two of the methods define distance measures to compare typed texts. The third method examines feature distribution parameters. Some modification of one of the methods has been proposed. The methods have been tested and their performance compared on two data sets. Finally the three methods have been combined to generate another decision rule, which has been also compared with the three original ones.

1 Introduction

Keystroke dynamics is one of behavioral characteristics which may be applied to classify, authenticate or identify users. The advantage of the keystroke dynamics is that it is a very natural type of biometrics which does not require any special hardware. Moreover, it is not as intrusive as some other methods. It is possible to collect the keyboard rhythm information during the normal computer usage. The difficulty with behavioral features is that they are not stable along time. One's typing rhythm depends not only on the time when it is recorded, but also on the type of hardware used and the text being typed. That is why the applications based on keystroke dynamics usually give higher error rates than those analysing some physiological features like for example fingerprint, palm, iris etc. In spite of this keystroke rhythm analysis may be used as an additional method to protect a system.

The study presented in this paper is a part of a research project focused on creating a biometric security system for mobile workstations. It is being realized at the Department of Knowledge Engineering of the Faculty of Electronics, Telecommunications and Informatics at the Gdansk University of Technology. To protect mobile

Agata Kolakowska
Gdansk University of Technology
e-mail: agatakol@eti.pg.gda.pl

devices, the system will analyse both physiological and behavioral characteristics. One of the modules will perform the keystroke dynamics analysis to authenticate users.

2 Analysis of Keystroke Dynamics

Many approaches have been investigated in order to classify, authenticate or identify users on the basis of their keystroke dynamics. Some of them apply common classifiers trying to discriminate between the samples of an authorised user and all other samples [1, 2, 3]. Methods of another group, usually used in outlier detection, assume there are only samples of one user and they try to find a boundary surrounding this set of samples [4].

The type of input data is what different methods have in common. Usually, to collect training data, the times of pressing and depressing all or some of the keys are recorded. Such data let us extract features which are usually single key times, i.e. the times between pressing and depressing these keys; and sequences times, which are the times between pressing the first and pressing or depressing the last key of a sequence.

This study focuses on three methods. The first two are based on distance measures defined to estimate the similarity between two typed texts. The first measure, called relative, is the degree of disorder of an array. It has been proposed and thoroughly explored in [5, 6]. Let us suppose that two text samples are represented as ordered arrays A and B containing the times of sequences of a given length. Only the sequences which appear in both texts are taken into account. The distance between these two arrays is defined as:

$$d(A, B) = \frac{\sum_{i=1}^n d_i(A, B)}{d_{max}}, \quad (1)$$

where n is the length of table A and B ; $d_i(A, B)$ is the distance between the position of the i -th element in table A and its corresponding element in table B ; d_{max} is the maximum disorder which is $n/2$ if n is even or $(n^2 - 1)/2$ otherwise. The distance falls within the range $[0, 1]$.

The second measure, called absolute, is defined as follows:

$$d(A, B) = 1 - s/n, \quad (2)$$

where s is the number of similar sequences. Two sequences with their duration times t_1 and t_2 are regarded as similar if

$$1 < \max(t_1, t_2) / \min(t_1, t_2) \leq t, \quad (3)$$

where $t > 1$ is a parameter which has to be adjusted experimentally.

To classify a new sample X a simple minimum distance rule may be applied, i.e. X is classified as a sample of user A if the mean distance between X and samples

belonging to A is the smallest. In the case of authentication the decision rule becomes more complicated. Suppose a new sample X appears and it claims to be a sample of user A . First of all the mean distance between A 's samples $m(A)$ has to be calculated. Then X is identified as belonging to the user A if

1. the mean distance between X and A 's samples $md(X, A)$ is the smallest and
2. X is sufficiently close to A , i.e. $md(A, X)$ is smaller than the average distance between A 's samples $m(A)$ or $md(A, X)$ is closer to $m(A)$ than to any other $md(B, X)$.

The rules above may be presented as the following inequality:

$$md(A, X) < m(A) + k[md(B, X) - m(A)], \quad (4)$$

where the parameter $k \leq 1$ enables to balance between the false acceptance rate (FAR) and the false rejection rate (FRR). The higher the value of k , the lower the rejection rate. If $k = 1$ the above inequality describes the classification rule. Lower values of k mean more secure system.

The authors of [5, 6] proposed some modifications of the original distance measures (1) and (2). If one wants to take into account the duration times of sequences of different lengths it is possible to sum the distances. For example to take advantage of both 2-key and 3-key sequences the distance measure may be defined as:

$$d(A, B) = d_2(A, B) + d_3(A, B) * n_3/n_2, \quad (5)$$

where $d_2(A, B)$ and $d_3(A, B)$ are the distances based on sequences of length 2 and 3 respectively; n_2 and n_3 are the numbers of these sequences.

In this study a different way of incorporating sequences of different lengths has been introduced. The idea is to represent a text by one ordered array which contains the times of sequences of different lengths. Usually it would give the same results as (5). The difference may appear in case of relative distance (1) when the time of a sequence of length l is shorter than the time of sequence of length m for $l > m$. However this happens very seldom, it is also worth noting. Suppose there are two texts A and B and there are only four common sequences: ab , $5c$, apa , wml . Let the times for these sequences for the two texts be:

1. $A : t(ab) = 0.150, t(5c) = 0.230, t(apa) = 0.250, t(wml) = 0.300;$
2. $B : t(ab) = 0.140, t(apa) = 0.230, t(5c) = 0.235, t(wml) = 0.290.$

The distance between A and B calculated according to (5) would be 0, because the order of 2-key elements in A and B is the same and the order of 3-key elements in the corresponding arrays is the same as well. Whereas the distance calculated after creating one array for each text would be greater than 0. This subtle difference might be meaningful if the number of common sequences is low.

The described methods have been compared with a method base on the analysis of feature distribution parameters. Let us suppose there are some texts of user A represented as feature vectors containing the duration times of sequences or single keys. Then for each feature its mean value μ_i and standard deviation σ_i may be

estimated. When a new sample X appears and it claims to be a sample of user A then the following inequality is checked for every feature:

$$\mu_i - \sigma_i < X_i < \mu_i + \sigma_i \quad (6)$$

where X_i is the i -th feature of X , $i = 1, 2, \dots, n$, n is the number of features. If the fraction of features which fall within the interval defined by (6) exceeds a pre-specified minimum value, then the user X is authorised as A .

3 Experiments

3.1 Training Data

A web application has been created to collect training data. When a user of this application typed a text, the times of pressing and depressing all keys were saved in a file. Fourteen volunteers took part in the experiments. They were given one fixed text, which was a piece of a famous Polish poem and it contained about 600 characters. The volunteers were asked to retype the text at least four times. They could not write a text more than once a day. They were allowed to make corrections while typing in the way they preferred, i.e. using backspace, delete, moving with arrows or mouse. The preferences in correcting errors could also characterize users. However in these experiments they were not examined. Even if the texts were not rewritten perfectly, they were still taken into account. Finally the training set contained 58 texts written by 10 users. The number of texts per user was between 5 and 8. Moreover there was also an additional set of 26 texts written by other users, i.e. users not present in the main training set.

Two types of error rates were estimated. The false rejection rate (FRR) was calculated on the basis of 58 identity verifications of the users from the training set in a standard leave-one-out validation method. The false acceptance rate (FAR) consisted of two components. The first one was estimated using the examples of the 10 users in the main training set. Each example, together with other examples of the same user, was taken out from the set. Then, while claiming successively to be one of the remaining nine users, its identity was verified. It gave $58 \times 9 = 522$ tests. The next FAR component was estimated using the 26 examples from the additional set of examples. Each of these examples was tested as belonging to one of the 10 users present in the training set, which gave $26 \times 10 = 260$ trials.

The second data set was a set of free texts. The same volunteers were asked to type 7-10 lines on whatever they wanted to. The rules on the frequency of typing and making corrections were the same as in the case of the given text. In this way a training set containing 165 texts written by 11 users was created. Moreover there was another set containing 16 texts written by users not present in the training set. The error rates were estimated in the way mentioned above. The number of trials performed to calculate FRR was therefore 165, whereas the number of tests accomplished to calculate FAR was 1826 ($165 \times 10 + 16 \times 11$).

3.2 Experimental Results

The first experiments were done for the method based on the relative distance measure (1) called R-method. This method has been tested for the sequences of different lengths. First of all 2-key sequences common for all texts were found. Then the texts were presented as arrays containing the times of the 2-key sequences. As the number of all sequences of this length present in all the texts was very high, only the most frequent 20 sequences were taken into account. The method was tested for different values of the parameter k in (4). The results obtained in this case are not satisfying as it is shown in table 1, columns a). Then all 3-key sequences were taken into account. The number of those sequences was 72. The error rates were much lower now: FRR = 10.34% while FAR = 4.22%. There might be two reasons for this: greater number of features and longer sequences which mean more informative features. To verify these presumptions the method was also tested for 4-key and finally 5-key sequences. The numbers of them were 32 and 12 respectively. The results for 4-key sequences were better than for the 2-key ones, but slightly worse than for 3-key ones. The results for 5-character sequences were even worse than in the case of 2-key ones. The reason for this was that there were only 12 sequences of length 5. Noticing the influence of the number of features (i.e. sequences) on the error rates, it has been decided to combine sequences of different lengths in one feature vector. By combining sequences of lengths 2, 3, 4 and 5 in single feature vectors, 137-dimensional feature vectors were created. This approach brought about significant improvement in the performance of the method. The FRR became 6.90% while FAR 3.96% as it may be seen from table 1, columns c).

The R-method has been also tested for free texts. It is obvious that the number of common sequences in this case was much lower, because they were extracted from completely different texts. There were 25 2-key sequences and one 3-key sequence. All these 26 features were put together in one feature vector. The results obtained for these data are shown in columns a) of table 2. To reduce the error rates it has been decided to make the feature vectors longer by adding the times of pressing single keys. The 20 most frequent keys were selected and such 20 features were added to the previous feature vectors. As it may be seen from columns b) of table 2 the error rates became much lower: FRR = 3.65% while FAR = 2.74%.

The same experiments were performed for the method based on the absolute distance measure (2) called A-method. First of all the method has been tested to select the value of parameter t in (3). The best results were obtained for $t=1.25$ and $t=1.35$, so the value of 1.25 was chosen for experimenting with method A for all data sets. In the case of given text the experiments were done for the sequences of lengths from 2 to 5 and finally for feature vectors created by combining all the sequences. The best results were obtained for the combined data which is shown in table 3, columns c). In this case the FRR was 8.62% while FAR 6.65%.

In the case of free texts the A-method did not show the same characteristics as method R, i.e. adding the times of pressing single keys did not improve the performance. The best results for this set of data was obtained for feature vectors created by combining 2 and 3-key sequences. These results have been show in table 4.

Table 1 Error rates for the R-method obtained in authenticating users on the base of the given text

sequence length:	a) 2		b) 3		c) 2,3,4,5	
k	FRR[%]	FAR[%]	FRR[%]	FAR[%]	FRR[%]	FAR[%]
0.3	53.45	3.45	31.03	0.38	31.03	0.38
0.4	43.10	3.58	20.69	0.38	20.69	0.51
0.5	41.38	4.35	17.24	0.38	13.79	0.51
0.6	39.66	4.60	17.24	0.77	12.07	1.02
0.7	36.21	5.75	13.79	1.79	10.34	1.79
0.8	34.48	7.16	10.34	4.22	6.90	3.96
0.9	31.03	8.70	10.34	7.16	5.17	6.78
1.0	29.31	10.10	10.34	10.61	5.17	10.74

Table 2 Error rates for the R-method obtained in authenticating users on the base of free texts

sequence length:	a) 2, 3		b) 1, 2, 3	
k	FRR[%]	FAR[%]	FRR[%]	FAR[%]
0.3	36.97	0.99	18.18	0.55
0.4	32.12	1.31	12.12	0.77
0.5	28.48	2.08	7.88	0.99
0.6	24.24	3.12	7.27	1.92
0.7	23.03	4.05	3.64	2.74
0.8	21.82	5.20	1.21	4.33
0.9	21.82	7.23	1.21	6.52
1.0	20.61	9.80	1.21	9.86

Table 3 Error rates for the A-method obtained in authenticating users on the base of the given text

sequence length:	a) 2		b) 3		c) 2,3,4,5	
k	FRR[%]	FAR[%]	FRR[%]	FAR[%]	FRR[%]	FAR[%]
0.3	50.00	2.94	27.59	0.13	22.41	0.13
0.4	41.38	3.20	22.41	0.26	20.69	0.38
0.5	36.21	3.71	20.69	0.77	15.52	0.51
0.6	32.76	4.48	20.69	1.15	13.79	1.41
0.7	29.31	5.75	17.24	3.07	13.79	3.20
0.8	29.31	7.16	15.52	4.86	12.07	4.86
0.9	27.59	8.70	13.79	7.93	8.62	6.65
1.0	27.59	10.49	12.07	10.49	6.90	10.74

Table 4 Error rates for the A-method obtained in authenticating users on the base of free texts described as combined sequences of length 2 and 3

k	FRR[%]	FAR[%]
0.3	28.48	1.37
0.4	25.45	1.97
0.5	21.21	2.57
0.6	17.58	3.01
0.7	14.55	4.16
0.8	13.33	5.26
0.9	12.12	7.34
1.0	10.30	9.69

The third method, which has been tested, was the method based on the analysis of feature distribution parameters. It has been investigated for different values of the minimum fraction of features (denoted as p in the result tables) which should fulfil the inequality (6) to acknowledge a user to be an authorised one. In this case the experiments were performed for all features, i.e. the sequences of different lengths and the times of pressing single keys put together in single feature vectors. Therefore the number of features was 157 for the given text and 46 for free texts. The results for these two sets of data are presented in tables 5 and 6.

Changing the parameters of the methods causes changes of the FRR and FAR values. The best way to compare different methods is to fix the FRR for example and then to adjust the parameters of the methods to achieve this value of FRR for every method. Then the methods may be compared and estimated on the basis of the FAR value obtained for the given FRR. Taking this into consideration and analysing

Table 5 Error rates for the method based on distribution parameters obtained in authenticating users on the base of the given text described as sequences of lengths 1, 2, 3, 4 and 5

p	FRR[%]	FAR[%]
0.45	0.00	32.48
0.47	1.72	26.73
0.49	5.17	20.84
0.51	5.17	16.11
0.53	10.34	12.92
0.55	13.79	10.49
0.57	18.97	7.93
0.59	27.59	5.50
0.61	34.48	4.22
0.63	44.83	3.20

Table 6 Error rates for the method based on distribution parameters obtained in authenticating users on the base of free texts described as sequences of lengths 1, 2, 3

p	FRR[%]	FAR[%]
0.47	1.21	38.77
0.49	2.42	32.15
0.51	4.85	25.96
0.53	6.06	20.32
0.55	6.06	16.59
0.57	10.30	13.31
0.59	12.73	10.84
0.61	15.76	8.32
0.63	21.21	6.96
0.65	27.88	5.64
0.67	33.94	4.71
0.69	39.39	3.50
0.71	46.06	2.30

carefully the presented results it may be noticed that for both data sets the method based on the feature parameter distributions gives the highest error rates. For example for FRR = 5.17%, the FAR is 6.78% for the R-method and 16.11% for the feature distribution method. When comparing the R-method and the A-method it may be shown that method R gives better results. For example when FAR is about 6.7% then the FRR is 5.17% for method R and 8.62% for method A. These conclusions are true for the two sets of data explored in this study, but they should not be generalised. To generate more general conclusions, more data sets should be investigated.

One more experiment has been performed in the efforts to improve the system's accuracy. The idea was to combine the three decision rules by voting. The combined rules were generated by method R, method A and the method based on feature distribution parameters. If more than one of the three rules classified the unknown sample as an authorised one then the final decision rule returned "authorised" as well, otherwise the final decision was "unauthorised". It was possible to reduce the error rate for the data set made of the given text. Eventually the FAR was 4.09% for FRR 5.17%, while for the best of the three combined methods used alone the FAR was 6.78% for the same value of FRR. Combining the rules did not improve the accuracy in the case of free texts data set.

4 Conclusions

In this study the performance of three known methods used to authenticate users on the basis of their keystroke dynamics was compared. The experiments were

performed on two sets of data: one created by retyping a given text and the other by typing any texts. The first two methods were slightly modified by merging the feature vectors representing the times of key sequences of different lengths. Moreover even the single key duration times were added to the same vectors. This modification improved the quality of the methods when compared to their original form. Greater number of common sequences in the texts usually gave better results. More improvement has been achieved after creating a voting decision rule which was a combination of the three rules tested separately.

Further experiments are planned to continue this study, including testing a few more methods, testing more data sets, gathering more training data for each data set. Finally a method is going to be chosen to be implemented in a real life environment as an additional biometric protection method for mobile workstations.

Acknowledgement. This work has been supported by the Polish Ministry of Science and Higher Education as the research project No. N N516 36793.

References

- [1] Monrose, F., Rubin, A.: Authentication via keystroke dynamics. In: Proceedings of the 4th ACM Conf. on Computer and Communication Security. ACM Press, New York (1997)
- [2] Curtin, M., Tappert, C., Villani, M., Ngo, G., Simone, J., Fort, H.S., Cha, S.H.: Keystroke biometric recognition on long-text input: a feasibility study. In: Proceedings of Student/Faculty Research Day, CSIS, Pace University (2006)
- [3] Clarke, N.L., Furnell, S.M.: Authenticating mobile phone users using keystroke analysis. *Int. J. Inf. Security* 6 (2007)
- [4] Tax, D.M.J.: One-class classification. PhD dissertation, Delft University of Technology (2001)
- [5] Bergadano, F., Gunetti, G., Picardi, C.: User authentication through keystroke dynamics. *ACM Trans. on Information and System Security* 5(4), 367–397 (2002)
- [6] Gunetti, G., Picardi, C.: Keystroke analysis of free text. *ACM Trans. on Information and System Security* 8(3), 312–347 (2005)

Determining Feature Points for Classification of Vehicles

Wieslaw Pamula

Abstract. The paper presents a discussion of feature point detectors characteristics for use in road traffic vehicle classifiers. Prerequisites for classifying vehicles in road traffic scenes are formulated. Detectors requiring the least computational resources suitable for hardware implementation are investigated. A novel rank based morphological detector is proposed which provides good performance in corner detection which is of paramount importance in calculating projective invariants of moving objects.

1 Introduction

Gathering information on road traffic flow is now commonly done with devices based on image processing [1], [2]. Sophisticated traffic control schemes, especially coordination of traffic flow in road networks, require not only simple data on the number of moving vehicles but also the parameters describing the dynamics of their movement. A good indication of movement dynamics is provided by distinguishing vehicle classes.

The knowledge of the size and type, in short class, of vehicles enables the estimation of flow dispersion parameters, traffic lane occupancy factors, queuing process parameters. Additionally discriminating objects, other than vehicles is also of much interest, specially for ensuring safety in interaction between for instance: pedestrians and vehicles, in traffic situations.

Currently available vision systems for traffic control evaluate the occupancy of detection fields for the assesment of vehicle class. Elaborate detection field configurations, mimicing vehicle shapes, enable rudimentary classification of vehicles in most cases inadequate for traffic control.

Wieslaw Pamula

Silesian University of Technology, Faculty of Transport,
Krasniskiego 8, 40-019 Katowice, Poland
e-mail: wieslaw.pamula@polsl.pl

The problem of vehicle classification requires resolving such issues as: definition of vehicle classes, aspects of world scene to camera image projection, complexity of discrimination algorithms, processing resources for implementing devised actions.

The paper is divided into four parts. First the prerequisites for vehicle classification are presented. The following section proposes a classification scheme based on geometric invariants. The scheme relies on robust acquisition of feature points describing geometric properties of vehicles. The third section introduces a novel feature detector based on morphological operations. The performance of the devised detector is compared with with feature points detectors of comparable computational complexity that is SUSAN and FAST. Conclusions are drawn in the last section

2 Vehicle Imaging

Traffic is monitored with CCTV cameras situated above or beside traffic lanes. This leads to a distorted view of travelling vehicles. Vehicle sizes are not preserved as they proceed along observed traffic lanes. To determine a space transform system for proper classification of objects a camera model is utilized. Using this model the field of view of the camera is calibrated to obtain the correct vehicle size estimation.

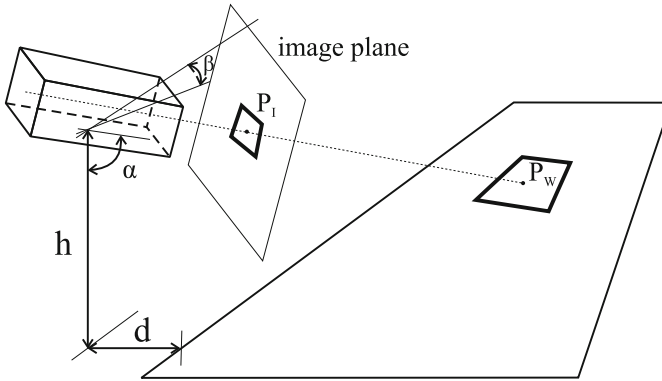


Fig. 1 Camera field of view.

Fig. 1. shows the position of the camera relative to the traffic lane. A camera is assumed located at a height h above the ground and a perpendicular distance d from the edge of the lane. It is directed at an angle β with respect to the road axis and tilted α to the ground.

The projection of a point in the coordinate system of moving vehicles onto an image plane of the sensor of the camera is additionally determined by hardware factors such as focal length of used optics, aspect ratio of the integrated light sensor [6]. The camera view model can be represented by the projection matrix M using

a homogeneous coordinate system in which the space point $P_W[X, Y, Z, 1]^T$ and the projected image point $P_I[x, y, 1]^T$ are related:

$$\begin{aligned} P_I &= MP_W \\ M &= HC \end{aligned} \tag{1}$$

Decomposing M into a hardware factors matrix H and a relative camera position matrix C , shows the transform details:

$$H = \begin{bmatrix} fS_x & 0 & p_x \\ 0 & fS_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} \mathbf{r}_1 & t_x \\ \mathbf{r}_2 & t_y \\ \mathbf{r}_3 & t_z \end{bmatrix} \tag{2}$$

where f is the focal length, S_x and S_y are the scaling factors, p_x and p_y are the offset parameters, $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ represent the components of camera rotation and are functions of α, β angles while t_x, t_y , and t_z are the translation parameters related to h and d .

Depending on values of these factors the projection distortion can be very significant.

Vehicle classes

Classifying vehicles for estimating traffic flow dynamics is dependent on the model of traffic which is utilized for carrying out control and planning schemes in ITS networks [3]. Number of classes is derived on the basis of required accuracy of modelling. It is usually between 3 and 8 the most common being 4. Considering a rough approximation of traffic flow model based on cellular automata four vehicle classes are adequate [4]. Fig. 2 presents models of vehicle classes used in traffic modelling. Vehicle shapes are modelled using polyhedrons which is an important characteristic for devising a classification scheme based on geometric invariants. Places of intersection of polyhedron side edges are point features so called corner points.

The aim of classification is to match an object to one from a set of class templates. A template may be defined in a number of ways: by a set of features, using a model, with a transform invariant. Matching is performed by analyzing the distance of the object description to the template. The distance measure is defined in the appropriate feature, model or transform space.

Feature based classification

Features such as corners, edge segments, patches of uniform pixel values and fork shapes are utilized for characterizing objects. Classes are defined by sets of relative positions of chosen features. A measure of matching distance for the prepared set is devised which is optimized for implementation. Successful classification requires a careful choice of features characterizing objects.



Fig. 2 Vehicle classes. Visible polyhedron sides marked

Objects moving in the field of view of the camera change their dimensions and appearance. A correct classification must use movement invariant features or somewhat limit the area of movement to minimize distortion.

Model fitting

Model fitting is based on matching wire models of objects. The models represent average object contours. Matching consists of finding models that coincide with object edges or segment approximations of object contours. The wire models are prepared off-line by projecting 3D polyhedrals as in fig.2 onto the image plane using eq.1 when the camera parameters are known or utilizing a field of view calibration program. When the field of view of the camera covers a large area and object features are hard to discern due to low camera resolution applying model fitting delivers a solution for classification. This approach is especially useful when tracking objects with occlusions in highly cluttered environment.

The resulting models are cut into segments, hidden segments are eliminated. A set of segments for each class is defined. The segments are additionally characterized by weights indicating their significance in total length of segments of the model. The matching process should also take into account the contents of interior of the model to avoid erroneous matching to road features instead of vehicles.

3 Geometric Invariants for Vehicle Classification

The imaging process of camera may be modelled by projective transform as shown in section 2. It does not preserve distances nor ratios of distances, a measure that is preserved is known as cross-ratio. Cross ratio is defined using points, lines or planes [6]. No invariant can be found for 3D points in general configuration from a single image [9]. Additional information on the geometric relations of the points is required. Using 6 points on two adjacent planes is adequate for defining an invariant [5]. Two of the points lie on the intersection line of the planes. Fig 3. illustrates the

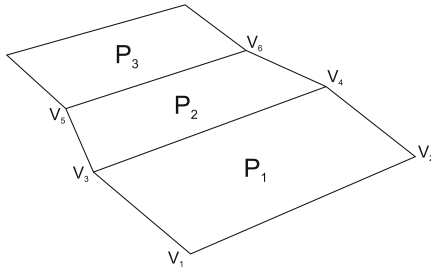


Fig. 3 Coplanar points for invariant calculation

invariant construction. Reviewing fig. 2 one observes multiple "plane" intersections that is common segments of polyhedron sides mapping vehicle shapes.

$$C_r(v_1, v_2, v_3, v_4) = \frac{d_{13}d_{24}}{d_{14}d_{23}} \tag{3}$$

where: v_i - plane points polyhedron side vertices, d_{ij} - distances between vertices v_i and v_j .

Using this measure an invariant is derived suitable for characterizing rigid polyhedral objects moving or variously placed in the field of view of a camera.

$$I = \frac{T_{123}T_{456}}{T_{124}T_{356}} \tag{4}$$

Where T_{ijk} are areas of triangles with vertices i, j, k . Expressing areas in terms of vertex coordinates on the image plane the invariant:

$$I = \frac{\begin{vmatrix} v_1 & 1 \\ v_2 & 1 \\ v_3 & 1 \end{vmatrix} \begin{vmatrix} v_4 & 1 \\ v_5 & 1 \\ v_6 & 1 \end{vmatrix}}{\begin{vmatrix} v_1 & 1 \\ v_2 & 1 \\ v_3 & 1 \end{vmatrix} \begin{vmatrix} v_4 & 1 \\ v_5 & 1 \\ v_6 & 1 \end{vmatrix}} \tag{5}$$

An object class is defined by a set of vertices selected to cover characteristic points which distinguish the class objects. These points must lie on two intersecting planes. Movement invariants for objects are sensitive to vertex coordinate extraction errors. It is vital to diminish transform and calculation errors.

It can be noted that a way to diminish errors is to define reference vertex points which lie as far apart as possible on the examined object. Fig. 2 shows preferable vertex points on vehicle models. These are roof top, front window vertices, head light centres. In order to enhance the robustness of classification a model is defined by two to four sets of reference points.

The off-line preparation of invariant values does not require the knowledge of camera parameters. 3D line drawings of models are adequate for estimating invariant values.

4 Detection of Corners

The search for corners - vertex points is effectively done using template matching, filtering in frequency or space domain. Template matching done with matrix operations requires considerable processing power, whilst filtering is usually performed with DSP devices.

Detectors such as Harris, DoG, SIFT, SURF evaluate space characteristics in the vicinity of a pixel at one or a number of space scales determining values of feature vectors attributed to this pixel. Their computational complexity is high thus impairing their use in hardware based vision systems suitable for road traffic applications.

Low complexity corner detectors such as SUSAN, FAST and based on morphological operations HMT promise attractive performance but require careful tuning of their parameters.

4.1 Rank Order Hit-or-Miss Transform

Morphological operations on gray scale images evaluate extrema of pixel values in image patches bounded by structuring elements. Templates are extracted using hit-or-miss transforms. The transform is defined using a number of descriptions [10] The prevailing approach is to use an interval operator which distinguishes pixels p of a range of values covered by a pair of flat structuring elements $SE1, SE2$:

$$\eta_{[v,w]}(F)(p) = \begin{cases} (F \ominus SE1)(p) & \text{if } (F \ominus SE1)(p) \\ & \geq (F \oplus SE2)(p) + t_1 - t_2 \neq \infty \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

The structuring elements in gray scale correspond to binary SEs: $SE1$ covers the object pixels, $SE2$ covers the background. Value $t_1 + t_2$ defines the range of values of the SEs.

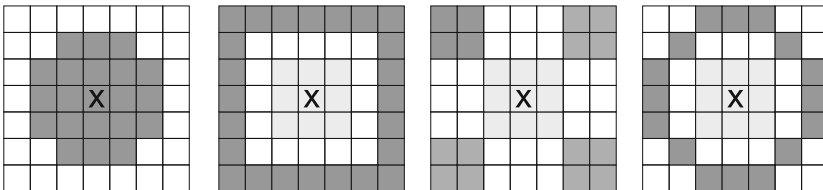


Fig. 4 Structuring elements for HMT: light gray - SE1, dark gray- SE2



Fig. 5 Vertex points on observed vehicles

The crucial task is the determination of the shape of the structuring element. The structuring element must accommodate for different positions of corners. Its size should match the size of objects in the observation field of the camera. Fig. 4 shows SEs which were examined for detection of corners in images. The structures detect groups of pixels which differ exactly by a set value from their neighbourhood companions, these are potential candidates for vertex points.

Corner detection tests carried out using sequences of road traffic images with different ambient lighting was erratic. The failure may be accounted to the high precision of template extraction. Discriminating conditions eq.6 had to be met exactly.

A way of weakening the conditions was to depart from the basic definition of erosion and dilation for gray scale images [11]. Fuzzy morphological operations and rank filters were considered.

Using order statistic filtering the set of pixel values covered by SE is sorted in ascending order and the k th value is chosen as the filter output. Minimum is rank 1 filter, maximum is rank n filter, n is the number of SE elements.

Rank filters enable partial classification of pixels covered by SEs. This brings insensitivity to corner shape and light changes. Tests were carried out using different



Fig. 6 Vertex points detected using FAST and SUSAN detectors

rank filters. The most successful proved to be the median filter. Fig.5. shows exemplary, out of a set of a few hundred traffic image sequences, results of corner detection using the devised median HMT detector. The detector used the right pair of structuring elements fig.4. First and second pair produced vehicle edges.

The detector was tested with different threshold values t_1 , t_2 . Results prove good detection of corner points. The detected points can be used to calculate geometric invariants of vehicles for distinguishing vehicle classes. Fig.6. presents corner detection using SUSAN and FAST detectors for comparison.

5 Conclusions

Rank order HMT based feature point detector extracts vehicle corners suitable for calculating geometric invariants. It is based on a sorting scheme which can be very efficiently implemented using a pipelined sorting network. Work is being done in implementing the devised detector in a FPGA based vehicle classifier for use in road traffic control.

References

- [1] Autoscope Terra, Data sheet. Image Sensing Systems USA (2010)
- [2] Traficam vehicle presence sensor. Traficon N.V. (2010)
- [3] Bretherton, R.D., Bodger, M., Baber, N.: SCOOT - Managing Congestion Communications and Control. In: Proceedings of ITS World Congress, San Francisco, pp. 678–701 (2005)
- [4] Marivoet, S., De Moor, B.: Cellular automata models of road traffic. *Physics Reports* 419, 1–64 (2005)
- [5] Song, S.B., Lee, K.M., Lee, S.U.: Model based object recognition using geometric invariants of points and lines. *Computer Vision and Image Understanding* 84, 361–383 (2001)
- [6] Zhu, Y., Senevirante, L.D., Earles, S.W.E.: New algorithm for calculating an invariant of 3D point sets from single view. *Image and Vision Computing* 14, 179–188 (1996)
- [7] Smith, S.M., Brady, J.M.: SUSAN – A new approach to low level image processing. *Int. J. Comput. Vis.* 23, 45–78 (1997)
- [8] Rosten, E., Drummond, T.: Machine Learning for High-Speed Corner Detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
- [9] Burns, J.B., Weiss, R.S., Riseman, E.M.: View variation of point-set and line-segment features. *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 51–68 (1993)
- [10] Naegel, B., Passat, N., Ronse, B.: Grey-level hit-or-miss transforms Part I: Unified theory. *Pattern Recognition* 40, 635–647 (2007)
- [11] Soille, P.: On morphological operators based on rank filters. *Pattern Recognition* 35, 527–535 (2002)

Edge Detection Using CNN for the Localization of Non-standard License Plate

L.N.P. Boggavarapu, R.S. Vaddi, H.D. Vankayalapati, and K.R. Anne

Abstract. Automatic license plate localization is one of the crucial steps for any intelligent transportation system. The location of the license plate is not same for all types of vehicles and in some developing countries the size of the license plate also varies. The localization of such a license plate requires a series of complex image processing steps in which edge detection plays major role, as the edges give crucial information regarding the location. In order to localize the license plate in real time, the amount of data to be processed must be minimized at the stage where the edges are identified. In this work we proposed Cellular Neural Network based edge detection for real time computation of edges. The performance of the proposed method has been evaluated over the collected database of 100 images with license plates located at different location. Based on the experimental results, we noted that Cellular Neural Network based edge detection improves the performance in non standard license plate localization when compared with traditional edge detection approaches.

1 Introduction

In the recent years due to the development in growth potential of the economies of the world and increase in purchasing power of the people towards two and four wheelers, the importance of traffic monitoring has raised the attention to Intelligent Transport Systems (ITS). These systems are to improve transportation safety and to enhance productivity through the use of advanced technologies. ITS tools are based on three core features; information, communication and integration that help

L.N.P. Boggavarapu · R.S. Vaddi · K.R. Anne

Department of Information Technology, V R Siddhartha Engineering College,
Vijayawada, India

H.D. Vankayalapati

Department of Computer Science & Engineering, V R Siddhartha Engineering College,
Vijayawada, India

operators and travelers make better and coordinated decisions. One of the most important applications of ITS is the Automatic license plate recognition (ALPR). ALPR is a mass surveillance system that captures the image of vehicles and recognizes their license number. This system has been utilized in many facilities, such as parking lots, security control of restricted areas, traffic law enforcement, congestion pricing, and automatic toll collection. License plate localization (LPL) or license plate detection is used to locate license plates from vehicle images accurately and efficiently [1]. This task is most crucial and significant step towards development of a complete ALPR. The success of higher level processing, localizing the licence plate, depends more on how good the edges are detected because edges contain a major function of image information. There are mainly three criteria to improve methods of edge detection. The first and most obvious is low error rate, the second criterion is that the distance between the edge pixels as found by the detector and the actual edge is to be at a minimum and the third criterion is to have only one response to a single edge. The present paper analyzes the performance of licence plate localization by using the above criteria of edge detection algorithms.

License plate localization means recognizing location of license plate region from offline vehicular images. Several algorithms have been proposed for Plate localization. These can be classified into morphology based, vertical edge density methods, edge statistics, rank filters, histogram analysis, artificial neural networks, fuzzy and template based methods. A morphological based approach is used by [2] on modifying the conventional approach to yield a cleaner result on complex images by applying heuristics. The methods like edge based & Histogram are normally simple and fast but they are sensitive towards noise [3]. A method based on Rank filter is used by [3] for localization of license plate regions is not suitable for skewed license plates. The method based on vertical edge detection followed by binarization is applied to a Swedish license plates are reported by Bergenudd [4] will not give better results for non-uniformly illuminated plates. The method based on artificial neural network (ANN) for the localization of Chinese Vehicle License Plates is not suitable for more complex back ground plates. Another method Sliding concentric windows is used in [5] for faster detection of region of interest. The method based on Horizontal Sobel's operator for the localization of Spanish license plate in the low saturation areas of the image is for plates with white background and black characters [6]. The method based on vertical edge density methods is reliable when the blur is very low [7]. A method to localize Iranian license plate is done in [10]. This method uses Sobel's operator followed by Otsu's binarization method followed by morphological operations. A method to localize Saudi Arabian License plate is done by [11]. This approach reduces the computation time by detecting only vertical lines. Additional edge extraction based approaches are discussed by [12] and [13]. Even though several approaches are available to localize licence plate, in developing countries like India, there is a difficulty of localizing as the license plate is not affixed in its proper position.

In this work we analyze the performance of licence plate localization by using the said criteria of edge detection algorithms and exploits the use of CNN in real time computation of edges.

2 Proposed Method for Licence Plate Localization

The proposed methodology involves the step, isolation of the license plate region from the input image [13]. Flow chart of the methodology is given in Fig. 1. It has two main processes namely, pre-processing and morphological operations. In the steps of preprocessing, enhance the contrast using cumulative probability density function, convert the obtained image in binary format and on that apply efficient edge detection algorithm (Here we made a comparative study of Canny, Sobel, Pre-witt, LOG and Cellular Neural Network based edge detection methods) and finally perform the important morphological operations to localize license plate.

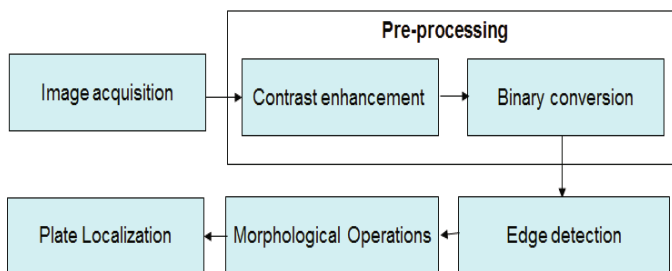


Fig. 1 Illustration of licence plate localization approach

3 Contrast Enhancement

Enhancing the contrast is very basic step in pre-processing because an image with good contrast detect the finest edges (The contrast enhancement technique). It is used to obtain an image within a given area, from a dark specimen and also improve visual quality of it [8]. It is an image processing technique that accentuates image detail by redistributing the range of color or gray display values.

At first step pixel matrix of the input image is calculated. The pixel matrix thus obtained has intensity values which are the discrete grey levels of the image. Let the image have L discrete grey levels ranging between X_0, X_1, \dots, X_{L-1} . Here X_0 is the minimum intensity and X_{L-1} is the maximum intensities of the image. $X(i, j) \in X_0, X_1, \dots, X_{L-1}$ for all (i, j) . Calculate the probability density of a grey level using $P(X_k) = n^{k/n}$ and cumulative density $C(X)$ of each grey level by using Equation (1).

$$C(X_k) = \sum_{j=0}^k P(X_j) \tag{1}$$

where $X_k = x$ for $k = 0, 1, \dots, L - 1$ Now apply transform function $f(x)$ for the input image, X by using the Equation (2).

$$f(x) = X_0 + (X_{L-1} - X_0)C(x) \tag{2}$$

4 Edge Detection

The contrast enhanced image is then converted into binary image (a digital image that has only two possible values for each pixel and is stored as a single bit (0 or 1) and this image is given as input to analyze the quality of edges from the edge detection algorithms. For fast and efficient processing of the images for localization, only most important information has to be processed, identifying and locating sharp discontinuities, and the rest has to be left out. It is one of the basic and very important image processing tasks [7]. In general, edge detection methods may be grouped into two categories: Gradient (Sobel, Prewitt), Laplacian (Marr-Hildreth, Canny). The Gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image. The Laplacian method searches for zero crossings in the second derivative of the image to find edges. Also there exists a new method Cellular Neural Networks (CNN) to find the accurate edges. The CNN model is a class of Differential Equation that has been known to have many application areas and high operational speed and uses the synaptic weight & templates to detect the discontinuity. This section of the paper is dedicated for the comparative study [14] of these edge detection techniques.

4.1 Gradient (Sobel and Prewitt)

An image can be understood as a matrix with x rows and y columns. It is represented with function say $f(x, y)$ of intensity values for each color over a 2D plane. Edges in an image can be detected using a periodical convolution of the function f with specific types of matrices m is shown in Equation (3) [3].

$$f'(x, y) = f(x, y) * m[x, y] = \sum_{i=0}^{w-1} \sum_{j=0}^{h-1} f(x, y) m[\text{mod}_w(x - i), \text{mod}_h(y - j)] \quad (3)$$

where w and h are dimensions of the image represented by the function f and $m[x, y]$ represents the element in x^{th} column and y^{th} row of matrix m , is also called as convolution matrix. The convolution matrix defines how the specific pixel is affected by neighboring pixels in the process of convolution. The pixel represented by the cell y in the destination image is affected by the pixels $x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ according to the Equation (4).

$$y = x_0m_0 + m_0x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + m_6x_6 + m_7x_7 + m_8x_8 \quad (4)$$

The process of convolution is explained in the Fig. 2. For Sobel and Prewitt methods there are separate and standard convolution matrices given respectively.

$$G_x = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \quad G_y = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}; \quad G_x = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad G_y = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}$$

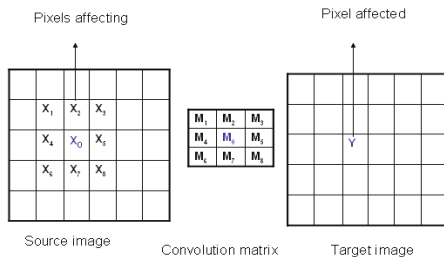


Fig. 2 Convolution process for edge detection using Sobel

Edge detection of the input image by applying Sobel and Prewitt is shown in Fig. 3

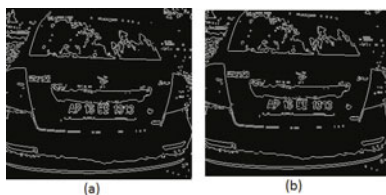


Fig. 3 (a)Edge detection by Sobel (b)Edge detection by Prewitt

4.2 Laplacian

The Marr-Hildreth Edge Detector is one of the edge detection methods can be seen as the application of a smoothing Filter followed by a derivative operation [15]. It is also known as Laplacian of a Gaussian function and is referred to as LOG. The general steps for the LOG are as follows: Step-1: Smoothing reduces the amount of error found due to noise. The smoothing is performed by a convolution with a Gaussian function Step-2: Apply a two dimensional Laplacian to the image: This Laplacian will be rotation invariant and is often called the "Mexican Hat operator" because of its shape. This operation is the equivalent of taking the second derivative of the image $\nabla^2 h_{\sigma}(u, v)$.

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{5}$$

Step-3: Loop through every pixel in the Laplacian of the smoothed image and look for sign changes. If there is a sign change and the slope across this sign change is greater than threshold, mark this pixel as an edge. Image after LOG is as shown in Fig. 4



Fig. 4 Edge detection by LOG

4.3 Canny Edge Detection

Canny edge detection algorithm is one of the most popular and optimal, runs in the following five steps: Step-1: The smoothing process is carried out by the Gaussian filter. Step-2: It is important to know about where the intensity of the image changes more. It is referred to as gradient of that image. The G_x and G_y are respectively as shown in Equation (6).

$$G_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad G_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \quad (6)$$

The magnitude, or edge strength, of the gradient is then approximated using $|G| = |G_x| + |G_y|$

Step-3: The direction of the edge is computed using the gradient in the x and y directions [16]. Step-4: Once the edge directions are found, non-maximum suppressions have to be applied. These are used to trace along the edge in the edge direction and suppress any pixel value (sets it equal to 0) that is not considered to be an edge. This will give a thin line in the output image. Step-5: Finally, Hysteresis is used to track along the remaining pixels that have not been suppressed. Hysteresis uses two thresholds and if the magnitude is below the first threshold, it is set to zero (made a nonedge). If the magnitude is above the high threshold, it is made an edge. And if the magnitude is between the 2 thresholds, then it is set to zero unless there is a path from this pixel to a pixel. Edge detection after completion of above 5 steps using Canny operator is shown in Fig. 5.



Fig. 5 Edge detection by Canny

4.4 Edge Detection by Cellular Neural Networks (CNN)

In this subsection, edge detection is done by using Cellular Neural Networks (CNN) [17]. The basic building block in the CNN model is the cell. The CNN model consists of regularly spaced array of cells. It can be identified as the combination of cellular automata and neural networks [18]. The adjacent cells communicate directly through their nearest neighbors and other cells communicate indirectly, because of the propagation effects in the model. The original idea was to use an array of simple, non-linearly coupled dynamic circuits to process, parallelly, large amounts of data in real time. The CNN processor is modeled by Equation (7) and Equation (8) with x_i, y_i and u_i as state, output and input variables respectively. The schematic model of a CNN cell is shown in Fig. 6. The coefficients $A(i, j)$, $B(i, j)$ values and synaptic weights completely define the behavior of the network with given input and initial conditions. These values are called the templates [18]. For the ease of representation, they can be represented as a matrix which is shown in Equation (9). We have three types of templates here. The first one is feed forward or control template, the second is feedback template and the third is bias. All these space invariant templates are called cloning templates. CNNs are particularly interesting because of their programmable nature i.e. changeable templates. These templates are expressed in the form of a matrix and are repeated in every neighborhood cell.

$$x_{ij} = -x_{i,j} + \sum_{c(j) \in N_r(i)} A_{ij}y_{ij} + B_{ij}u_{ij} + I \tag{7}$$

$$y_{ij} = 0.5 (|x_{ij+1}| - |x_{ij-1}|) \tag{8}$$

CNN based edge detection on binary image can be achieved by using the following template set [19] as given in Equation (9) and result is shown in Fig. 7

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad G_y = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix} \quad I = -1 \tag{9}$$

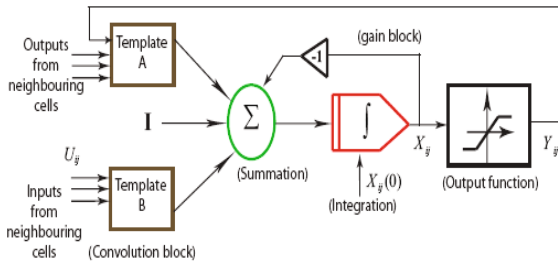


Fig. 6 Schematic Representation of a CNN cell



Fig. 7 Edge detection by CNN

5 Morphological Operations

Now we apply the Morphology Operations (MO) to the edge detected image. This is constructed on the basis of set and non linear theory for image analysis and processing. Now a days, the MO has been widely used in the fields of image processing and computer vision and uses structure element for measuring the morphology of image to understand and solve problems accurately. A structure element can be of different geometrical shape and size. But the shape of structure element [20] widely used now is the form of line (usually horizontal or vertical), and the size is usually computed according to the specific conditions. Dilation and erosion are two basic morphological processing operations which are considered by convolving the structure element with the image. They are defined in terms of more elementary set operations. Dilation is used to fill in the gaps or holes, but on the contrary, erosion is often used to remove irrelevant details from binary image. Let A be a set of pixels and let B be a structuring element. Let ${}^A B$ be the reflection of about its origin and followed by a shift by s . Dilation is denoted as $A \oplus B$ suppose that the structuring element is a 3×3 square, with the origin at its center. Dilation and erosion is defined mathematically using the Equation (10) and Equation (11) respectively.

$$A \oplus B = \{S / (({}^A B)_s \cap A) \subseteq A\} \quad (10)$$

$$A \ominus B = \{S / (B)_s \subseteq A\} \quad (11)$$

Closing is nothing but applying erosion on the dilation. In erosion, the value of the output pixel is the minimum value of all the pixels in the input pixel's neighborhood. In a binary image, if any of the pixels is set to 0, the output pixel is set to 0.

6 Evaluation of CNN Based Edge Detection over Licence Plate Localization

Edge detection is one of the important steps in license plate recognition. It is observed from the Fig. 1.8(b) that some parts having more number of edges likely to be suspected parts for having license plate but in fact all that parts are not license plate regions. Fig. 8 represents the recognized license plate and Fig. 1.8(c) represents the original license plate. If the edge detection image is not detecting more

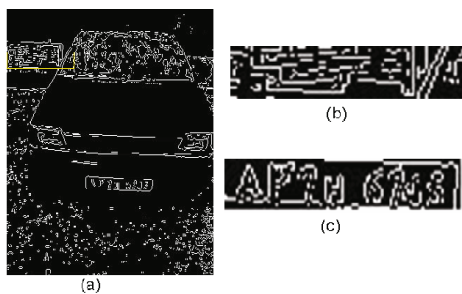


Fig. 8 (a) Edge detection output (b) Recognized license plate (c) Correct license plate

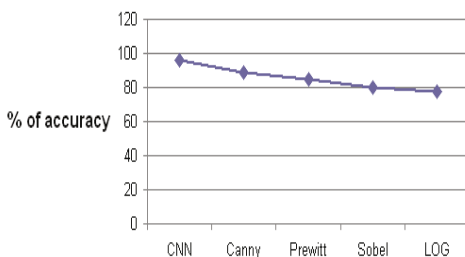


Fig. 9 Performance evaluation of the different edge detection methods versus % of accuracy

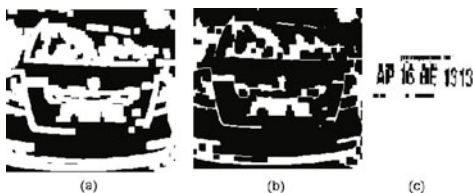


Fig. 10 Morphological operations performed over image after CNN based edge detection (a) Image after dilation (b) Image after close operation (c) Localized licence plate

edges, then it is a problem (some times license plate is also not recognized). This bifurcation can be possible exactly after applying CNN based edge detection.

From the experimental study, it is observed that among the explained edge detection methods in this paper, CNN has the capability of identifying the edges in the image accurately. Fig. 9 is the performance evaluation of the different edge detection methods. The said algorithm is tested using Lab View.

With this reason we have used CNN based edge detection and then apply the morphological operations. The morphological Operations of the image is shown in Fig. 10. Consider a window size 3.5 inch tall, 10 inches wide (Standard Indian license plate size) and find out the existence of license plate. The output obtained is the license plate that is located in the original image as shown in Fig. 1.10(c).

Acknowledgements. This work was supported by TIFAC - CORE (Technology Information Forecasting and Assessment Council - Center Of Relevance and Excellence) in the area of Telematics, a joint venture of DST (Department of Science and Technology), Govt. of India, V.R. Siddhartha Engineering College, Vijayawada, India.

References

- [1] Sun, G.: The Location and Recognition of Chinese Vehicle License Plates under Complex Backgrounds. *Journal of multimedia* 4 (2009)
- [2] Dubey, P.: Heuristic Approach for License Plate Detection, *IEEE Transactions*, September 15 (2005)
- [3] Martinsky, O.: Algorithmic and Mathematical Principles of Automatic Number Plate Recognition System, BRNO University of Technology (2007)
- [4] Bergenudd, E.: Low-Cost Real-Time License Plate Recognition for a Vehicle PC, *KTH Electrical Engineering* (2006)
- [5] Anagnostopoulos, C.N., Loumos, I., Kayafas, E.: A license plate recognition algorithm for Intelligent Transport applications. *IEEE Transactions on Intelligent Transportation Systems* 7(3) (2006)
- [6] Garcia-Osorio, C., Diez-Pastor, J.-F., Rodriguez, J.J., Maudes, J.: License Plate Number Recognition New Heuristics and a comparative study of classifier. In: 5th International Conference on Informatics in Control Automation and Robotics (2008)
- [7] Jain-Feng, X., Shao-Fa, L., Mian-Shu, Y.: Car License Plate Extraction using Color and Edge. In: *Int. Conf. Machine Learning and Cybernetics*, vol. 6. IEEE Press, Los Alamitos (2004)
- [8] Rattanathammawat, P., Chalidabhongse, T.H.: A Car Plate Detector using Edge Information. *ISCIT* (2006)
- [9] Al-Ghaili, A.M., Mashohor, S., Ismail, A., Ramli, A.R.: A New Vertical Edge Detection Algorithm and its Application. In: *Int. Conf. Computer Engineering Systems, ICCES* (March 2008)
- [10] Mahini, H., Kasaei, S., Dorri, F.: An Efficient Features-Based License Plate Localization Method. In: 18th International Conference on Pattern Recognition (2006)
- [11] Sarfraz, M., Ahmed, J., Ghazi, S.: Saudi Arabian License Plate Recognition System. In: *International Conference on Geometric Modeling and Graphics* (2003)
- [12] Ozbay, S., Ercelebi, E.: Automatic Vehicle Identification by Plate Recognition, *World Academy of Science, Engineering and Technology* (2005)
- [13] Shapir, Dimov, V., Bonchev, S.D., Velichkov, V., Gluhchev, G.: Adaptive License Plate Image Extraction (2006)
- [14] Juneja, M., Sandhu, P.S.: Performance Evaluation of Edge Detection Techniques for Images in Spatial Domain. *International Journal of Computer Theory and Engineering* 1 (2009)
- [15] Neoh, H.S., Hazanchuk, A.: Adaptive Edge Detection for Real-Time Video Processing using FPGAs. *Global Signal Processing* (2004)
- [16] Canny, J.: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (1986)
- [17] Chua, L., Roska, T.: Cellular neural networks and visual computing: foundations and applications. Cambridge University Press, Cambridge (2005)
- [18] Kozek, Chua, L.: Genetic Algorithm for CNN Template Learning. *IEEE Transactions on Circuits and Systems* 40 (1993)

- [19] Vamsi Prakash, M.: CNN based non-linear image processing for robust Pixel Vision. Alpen-Adria-University, Klagenfurt (2009)
- [20] Kasaei, S., Hamidreza: New Morphology-Based Method for Robust Iranian Car Plate Detection and Recognition. International Journal of Computer Theory and Engineering 2 (2010)
- [21] Paunwala, C.N., Patnaik, S.: A Novel Multiple License Plate Extraction Technique for Complex Background in Indian Traffic Conditions. International Journal of Image Processing, 4 (2010)

Artificial Neural Network Approach for Evaluation of Gas Sensor Array Responses

Iwona Rutkowska, Barbara Flisowska-Wiercik, Przemysław M. Szecówka,
and Andrzej Szczurek

Abstract. Neural processing was applied for evaluation of gas sensors responses. Thin layer type gas sensors based on TiO₂ sensing film were fabricated using sol-gel technology. Several variants of these devices were obtained by different technology parameters. They were characterized by exposition to air containing various volatile organic compounds. Neural processing was applied for classification of compounds according to the chemical function group and number of carbon atoms in molecules. Different statistics of correct classification obtained for the two variants deliver significant knowledge about sensing mechanisms and quality of information that may be extracted from sensor responses.

1 Introduction

Recent years have been characterized by a growing interest focused on measurement of air pollutants. It is caused by the social, economic and market impacts of the deterioration of the natural environment due to human activity. The ability to monitor and detect various gases is important for many applications. For example, there is a great interest in obtaining information about air quality. Rapid detection and quantification of chemical species are also important in optimization of industrial processes. Monitoring of gases, such as CO, O₂ and CO₂ at high temperatures

Iwona Rutkowska
Faculty of Chemistry
e-mail: iwona.rutkowska@pwr.wroc.pl

Barbara Flisowska-Wiercik · Andrzej Szczurek
Faculty of Environmental Engineering
e-mail: andrzej.szczurek@pwr.wroc.pl

Przemysław M. Szecówka
Faculty of Microsystem Electronics and Photonics,
Wrocław University of Technology, Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
e-mail: przemyslaw.szecowka@pwr.wroc.pl

with the opportunity for feedback control can lead to significant energy savings, as well as minimization of emissions across power, chemical, steel and other manufacturing industries. In working environment a lot of attention is focused on detecting the presence and concentration of toxic or otherwise dangerous gases that may come from spills and leaks. Another broad application area of the gas measurements is quality control and industrial monitoring, particularly in such industries as food processing, perfume, beverage and other chemical products.

Various kinds of methods and techniques are available today in environmental monitoring. They typically involve time-intensive collecting samples and analyzing them in an electrochemical analyzer, spectrophotometer, gas chromatograph or gas chromatograph-mass spectrometer. These instruments are relatively expensive and of significant size (too bulky for in situ operation). In addition, skilled personnel, careful maintenance and complex calibration processes are necessary. Consequently, the number of monitoring locations is strongly limited. For that reason, although conventional instruments work very well, many applications need devices that are smaller, more portable, cheaper and even disposable. This demand grows in many areas of human activity. Thus, a considerable amount of scientific research has been directed towards the development of reliable and selective solid-state gas sensors to reduce cost and increase the number of monitoring locations. Among sensors which exploit metal oxide semiconductors, metal oxide semiconducting field effect transistors (MOSFETs), quartz crystal microbalance (QCM), surface acoustic, those based on conductance changes of semiconducting metal oxides result to be the most suitable devices to fulfill a compromise between some principal advantages (low costs, high sensitivity, short response time, simple measurement electronics) and some well-known disadvantages (lack of selectivity, poor long-term stability, humidity dependence) [4]. Traditionally, chemical sensing methods primarily rely on the inherent selectivity of the sensor to obtain quantitative information or identify the presence or absence of analyte [6]. Unfortunately, in many cases, gas sensors cannot achieve the required selectivity. Many strategies may be used to solve this problem. One of them is based on a sensor array coupled with powerful pattern recognition methods.

In reality, metal oxide semiconductors respond to multiple gases. Consequently, a tested gas activates many of the sensors in the array. The fact that a sensor responds to more than one compound is often treated as a serious drawback. However, this feature can sometimes be useful. The basic idea in this strategy is to create sensing elements of the array which exhibit distinct but overlapping broad sensitivity profiles for the range of gases of interest. The sensor array used should generate a pattern of responses that are discernibly different for various samples.

The data collected from the sensor array is a powerful source of information about tested gas. However pattern recognition and multivariate analysis methods are usually required to extract the maximum useful information from the sensor output signals (sensor array response). The term pattern recognition is used to describe the methodology of solving feature extraction, classification and identification. The best known approaches for pattern recognition in sensor systems are: template matching, statistical classification, syntactic or structural matching and neural processing [5].

Among these techniques, artificial neural networks (ANNs) are generally considered as most promising tools for chemical sensing particularly in the case of systems based on a sensor array [7]. The sensing properties of sensor array depend upon many factors. For example, the individual sensors should respond to a broad range of compounds and should not be highly specific in their responses. The responses from the elements in an array to a specific gas should be as independent as possible (in the statistical sense) to maximize the information being gathered for that compound. Different approaches are used to perform this requirement. They exploit: chemical properties of the active materials, physical parameters of the sensing layer, surface modification, sensor design, physical and chemical modification of the sample (before it contacts with the chemical sensitive layer) and mode of sensor operation. In this paper our attention is focused on sensor array containing sensing elements produced in different technological conditions. It is shown that this array coupled with an artificial neural network can be used for the measurement of organic vapors in air.

2 Sensor Technology and Characterisation

Gas sensors consist of substrate, electrodes, sensing material and heater. The key component is sensing material, in this case TiO_2 film fabricated using a sol-gel technique [1, 2, 3]. The porous alumina (Al_2O_3) was used as substrate. This element was provided by a platinum heater on the back-side and gold interdigitated electrodes on the front-side, all screen-printed (Fig. 1). The heater permitted to maintain the sensing material at working temperature. Gold was selected as the electrode material because of chemical resistivity and stability at relatively high temperature. Electrodes were formed with a spacing of 1 mm. TiO_2 films were deposited onto substrate by dip-coating method. The substrates were kept immersed in the solution for two different times – 10 and 60 s. The substrate was withdrawn with a speed of 34 mm/min. The deposited layers were dried in atmosphere air for minimum 3 h. Then TiO_2 thin films were annealed in air for one hour and at temperature reaching 600°C with a controlled gradient.

The experimental set-up used for the characterization of sensors was composed of: an apparatus for the preparation and delivery of reference and measured gas, test chamber, power supply and electric current measurement unit. The first element consisted of apparatus for air purification, equipment for the preparation of gas samples, gas line (teflon tubes) and a diaphragm pump. The cartridges filled with silica gel and activated carbon were applied to dry and purify air. Gas samples were prepared by an evaporation method. In this case air containing vapors of alcohol, aldehyde or organic acid was chosen as gas mixture of interest. Desired amounts of these compounds were injected (using microsyringe) as a liquid into the heated coil and then vaporized in a stream of pure, dry air. The flow rate of air was precisely adjusted and controlled by a mass flow controller. The organic liquid was fully converted into the vapor. The sample was collected in a Tedlar bag. This method of sample preparation had the capability of generating different

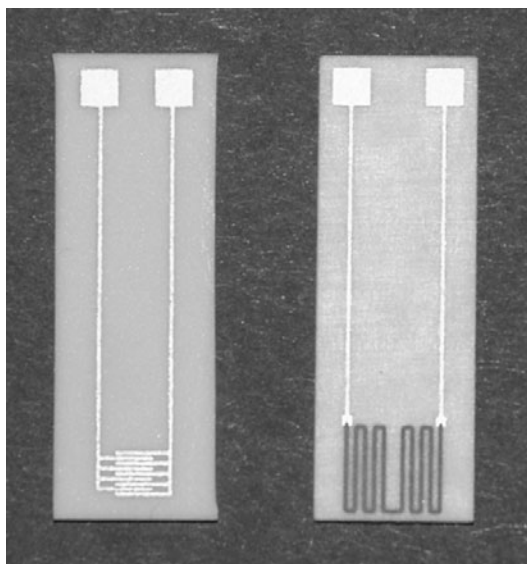


Fig. 1 Sensor structure - contacts, electrodes and resistive heater screenprinted on alumina substrate.

concentrations of VOCs in air. Concentration of these substances was determined by dosage, rate and time of air flow. Concentration of VOCs covers the range from tens to hundreds of ppm (part per million, volumetric).

The TiO_2 film fabricated on the substrate was mounted in an airtight glass test chamber. All measurements were performed keeping this compartment at room temperature. A test chamber was equipped with two ports for the gases (inlet and outlet) and a system of electrical connections. Sensing properties of TiO_2 films were evaluated using DC measurements. The constant voltage of 5V was applied between electrodes by using a stabilized power supply. A current intensity between two electrodes was measured by a picoammeter when the sensor achieved steady-state. A chemical composition of atmosphere inside this vessel was controlled. The test chamber was attached via plastic tubing to gas delivery system and a diaphragm pump. The heater of gas sensor was connected to voltage supplier. During experiments the sensors were heated by applying a voltage across the heating meander, whose resistivity value determined sensor's operating temperature. All electrical measurements presented in this work were carried out at the working temperature of 450 °C.

In our work, all tests were performed according to the same experimental procedure. Gas sensing properties of TiO_2 thin films were examined in dynamic conditions through the sequential exposures. The experiments involved exposing the sensor to a continuous stream of pure, dry air or gas sample. The measurement process consisted of three different phases. In the first phase, TiO_2 thin film was exposed to a reference gas – pure, dry air. In the second phase, the sensor was in the

stream of air containing target, organic species. This part of experiment was 8 min long. It was enough for sensor to reach a steady-state (a stable value of conductivity). The gas sample was introduced into a glass test chamber from the tedlar bag using the diaphragm pump. When the exposure was completed the third phase, with pure, dry air and no measurements, was activated. The main purpose of this phase was to clean the test chamber and return to the sensor baseline before the next measurement. The test chamber was purged for several minutes. During the measurement phases, air and test sample were allowed to flow at a fixed rate of 1 l/min. All operations in experiments were performed manually.

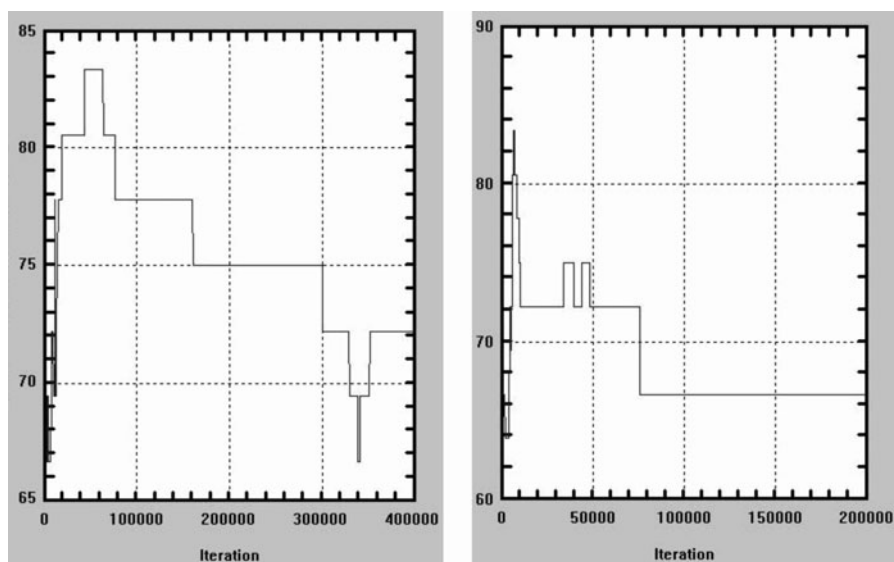


Fig. 2 Accuracy of neural networks applied for distinguishing between compound type - alcohol, aldehyde, acid, and its evolution during the training process; calculated for testing data set.

3 Neural Processing

Sensors capabilities were estimated in a series of experiments with neural processing. The performance of neural network – accuracy and generalization abilities, gained for the specific construction of training patterns, is quite good estimation of quality of information carried by the sensors. Eventually very practical information about sensors reaction for specific gases and capabilities for specific tasks may be revealed. In this case the input pattern for neural processing consisted of responses of 3 sensors for 9 volatile compounds: methanol, ethanol, propanol, formic aldehyde, acetic aldehyde, propionic aldehyde, formic acid, acetic acid and propionic acid. These compounds

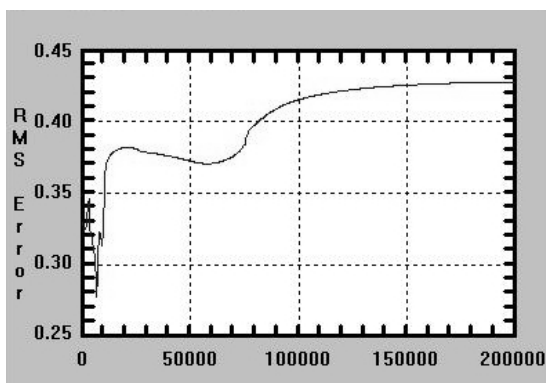


Fig. 3 Test set RMS error evolution. Neural network applied for distinguishing between compound type.

may be classified in two ways: according to a type (functional group) – alcohols, aldehydes, acids, and according to a number of carbon atoms in a molecule – 1, 2, 3 atoms. For each class, defined in each way there are 3 elements.

3.1 Classification According to Function Group of the Compound

In the first variant feedforward neural networks were trained to distinguish between the following 3 classes of compounds: class 1 – methanol, ethanol, propanol, class 2 – formic aldehyde, acetic aldehyde, propionic aldehyde and class 3 – formic acid, acetic acid, propionic acid.

The input vectors contained responses of 3 sensors to various concentrations of all the 9 compounds. Corresponding output vectors contained the classes, coded in binary way in 3 dedicated outputs. The data set was divided to two exclusive parts for training and testing. Various structures of neural networks were investigated, usually with 2 hidden layers. Two kinds of activation functions in neurons were applied – sigmoid and hyperbolic tangent. Depending on activation function, the output patterns were scaled to 0.15 and 0.85 or +/- 0.85. RMS error seems to be a classic and commonly recognized measure of neural network performance. This is not the case for sensor systems technology however. For metrological devices the more honest quality criterion is the worst case error, which says that the precision of e.g. gas concentration meter is "not lower than ...". It is yet different for classification problems, where the answers shall be perceived as either correct or wrong. Thus for this research the statistics of proper responses for the testing data set was selected as a measure of neural network (and system) performance. For output coding described above, responses below 0.5 (or below zero respectively) were counted as correct for the negative patterns whilst responses above 0.5 (or above zero) were counted as correct for the expected positive response. The responses not matching

these criteria were counted as misclassifications. The statistics were observed along the training process and the peak result was taken into account as a measure of the performance available to reach for the given set of sensors in the context of the current task. Series of multi-layer networks, with various architectures, were trained with classic BP algorithm supported by strong momentum and tiny noise added to input patterns. Sample results are presented in Fig. 2. The two plots show statistics of correct responses versus iterations for two neural networks. One of them contains 38 neurons in hidden layers, the other one – 52 neurons. Variation of accuracy during the training process is quite typical for this technique. In the begin the accuracy improves and then after reaching some top point it tends to decrease, due to overtraining. There are some exceptions of this generic scheme, partially caused by undeterministic behaviour of neural networks, partially because the statistics of correct responses is not exactly compatible with the RMS error. Classic RMS history is shown for reference in Fig. 3 (the plot shall be compared with the one presented in Fig. 2 on the right). In this case the factor of interest is not the final result but the highest performance reached during the training process. The top result obtained in several experiments reached 85 per cent of correct responses. Variations in number of neurons and activation functions selection (including hybrid constructions) didn't improve the result. There were significant differences in neural network performance, depending on method of division of the data set to training and testing. For random selection the results were substantially worse than for careful and systematic removing of patterns from training set. Thus it may be concluded that when applied for recognition of type of compound the sensors may provide correct classification in 85 % of cases, but no more. This result is not satisfying from pragmatic point of view. Such level of reliability is rarely accepted for gas sensor systems.

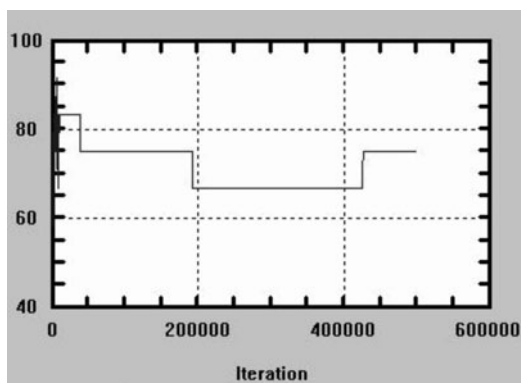


Fig. 4 Accuracy of neural network applied for distinguishing between alcohols and aldehydes, and its evolution during the training process; calculated for testing data set.

Better results were obtained for the recognition limited to alcohols and aldehydes. The training vectors were prepared in two variants. In each of them two dedicated binary outputs were applied to identification of compound (alcohol, aldehyde). Then the vectors containing responses for the acids were either marked as „00” (i.e. this is neither alcohol nor aldehyde – something else) or removed from the data set. The sample result is presented in Fig. 4. This time, for a very short period of training process the statistics reached 90 %.

3.2 Classification According to Number of Carbon Atoms

The second approach involved similar experiments focused on recognition of the compound classes defined in different way: class 1 – methanol, formic aldehyde, formic acid class 2 – ethanol, acetic aldehyde, acetic acid and class 3 – propanol, propionic aldehyde, propionic acid. Thus the key to classification was a number of carbon atoms in a compound molecule. For the same set of sensors responses, the output vectors were redefined properly, and the data was divided to training and testing parts again. Several structures of neural networks were investigated as previously. Sample results are presented in Fig. 5. Now the peak results of statistics are reaching 95-96%. This level of reliability is acceptable for some applications and commercial low cost devices. At this stage however, more important observation is better performance of sensors applied for classification performed according to a number of carbon atoms in a gas than for classification performed according to a compound type. The compatible RMS error evolution is shown in Fig. 6.

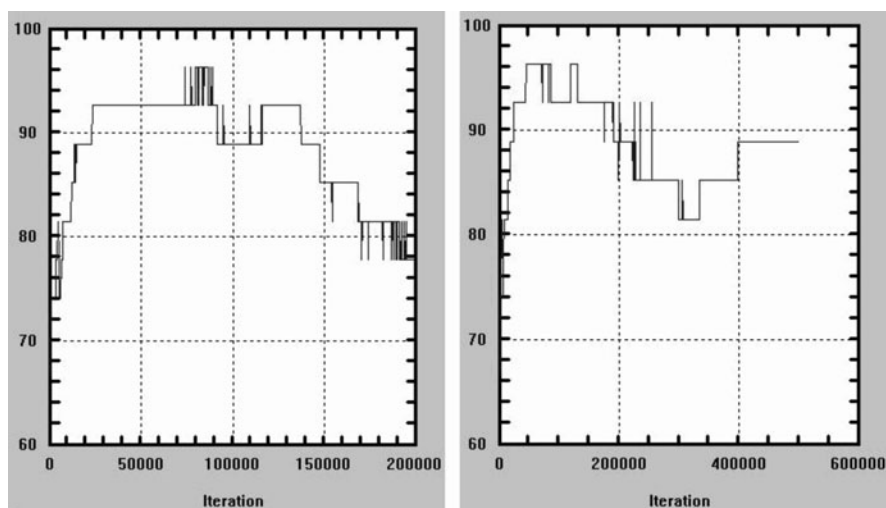


Fig. 5 Accuracy of neural network applied for distinguishing between compound series (1, 2, 3 atoms of carbon in a molecule), and its evolution during training process.

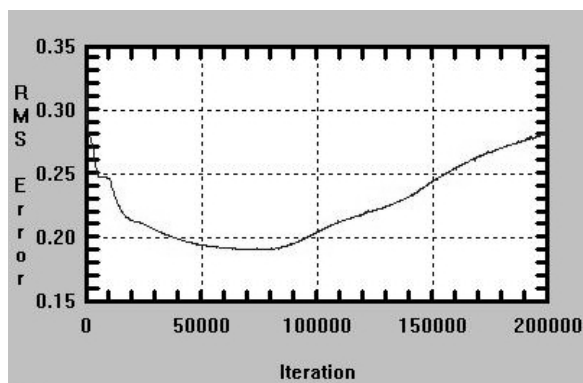


Fig. 6 Test set RMS error evolution. Neural network applied for distinguishing between compound series (1, 2, 3 atoms of carbon in a molecule).

4 Conclusions

Thin film gas sensors based on TiO_2 , fabricated by sol-gel technology were in-house developed and characterized. Poor selectivity of sensors was revealed, i.e. all the sensors react for all the compounds tested. This feature is quite common for semiconductor gas sensors. It is an obvious drawback but the last decade of gas sensor research have shown that it may be also advantage broadening the application field of sensors. Sensor array coupled with advanced processing of multidimensional data became the classic solution in this context. In this case we were quite close to prove that the same 3 sensors may be used to classify compounds in two completely different ways.

At the current stage it would be hard to apply the sensors in practice – for all the variants the misclassifications occur too frequently. Simultaneously the statistics of correct responses are high enough to believe that quite small improvement in the technology may lead to construction of sensors providing absolutely correct classification.

The key part of presented research however involved somewhat more original approach. The target was not the construction of a sensor system in strict sense but the estimation of quality of information delivered by the set of sensors. The trials were repeated several times, for varying neural network constructions to obtain more reliable results. The wide spectrum of sensors reactions for various compounds shows that mechanisms of conduction are sensitive for both the function group and number of carbon atoms in a molecule. It was shown however that in the latter case the statistics of correct classification was substantially higher. Thus the influence of carbon atoms is higher. This phenomena needs further explanation taking into account the details of sensing mechanism of TiO_2 thin layer. It is an interesting field

of further research. At this stage it suggests the proper direction of experiments with sensitive material preparation, sensor operating conditions and the prospective successful field of sensors application.

References

- [1] Kozłowska, K., Łukowiak, A., Szczurek, A., Dudek, K., Maruszewski, K.: Sol gel coatings for electrical gas sensors. *Optica Applicata* 35, 783–790 (2005)
- [2] Łukowiak, A., Kozłowska, K., Urbański, K., Szczurek, A., Dudek, K., Maruszewski, K.: The application of an artificial neural network in the processing of output signals from a gas sensor with sol-gel-derived TiO₂ film. *Materials Science - Poland* 25, 861–868 (2007)
- [3] Łukowiak, A., Maciejewska, M., Szczurek, A., Maruszewski, K.: Application of titania thin film for the discrimination between diesel fuel and heating oil. *Thin Solid Films* 515, 7005–7010 (2007)
- [4] Moseley, P.T., Norris, J., Williams, D.: *Techniques and mechanisms of gas sensing*. IOP Publishing Ltd., New York (1991)
- [5] Pardo, M., Kwong, L.G., Sberveglieri, G., Brubaker, K., Schneider, J.F., Penrose, W.R., Stetter, J.R.: Data analysis for a hybrid sensor array. *Sensors and Actuators B* 106, 136–143 (2005)
- [6] Sberveglieri, G.: *Gas sensors*. Kluwer Academic Publishers, Dordrecht (1992)
- [7] Shi, X., Wang, L., Kariuki, N., Luob, J., Zhongb, C.-J., Lua, S.: A multi-module artificial neural network approach to pattern recognition with optimized nanostructured sensor array. *Sensors and Actuators B* 117, 65–73 (2006)

Developing Decision Tree for Identification of Critical Voltage Control Areas

Robert A. Lis

Abstract. Assessing and mitigating problems associated with voltage security remains a critical concern for many power system operators. It is well understood that voltage stability, is driven by the balance of reactive power in a system. Of particular interest is to identify those areas in a system that may suffer reactive power deficiencies. Establishing the reactive power reserve requirements in these areas (VCAs), to ensure system integrity is of paramount importance. The approach is based on a Modal Analysis (MA) combined with Decision Trees (DTs). A database with user-friendly interface for storing/retrieving result of MA for each scenario was designed. The information in the database is used to construct DTs for each of the identified VCAs using key system attributes. In on-line application, the relevant attributes is extracted from a system snapshot and is dropped on DTs to determine which of the pre-determined VCAs can exist in the present system condition.

1 Introduction

Power system dynamic security assessment (DSA) is the process of determining the degree of risk in a power system's ability to survive imminent disturbances (contingencies) without interruption to customer service [7]. Power system security depends on the system operating condition as well as the contingent probability of disturbances, and therefore, DSA must consider these factors. Power system DSA has traditionally been conducted off-line using a variety of analytical techniques and the results used by human operators to guide real-time operation. However, the complexity of today's systems, and operating requirements, has largely rendered off-line DSA inadequate and a growing number of utilities are implementing on-line DSA systems [3] that can reduce, or eliminate entirely, the need for off-line analysis. In

Robert A. Lis

Institute of Electric Power Engineering,

Wroclaw University of Technology, Wybrzeze Wyspianskiego 27,

50-370 Wroclaw, Poland

e-mail: robert.lis@pwr.wroc.pl

this approach, a snapshot of the real-time system is captured and, using a variety of methods, the security is assessed for a large set of probable contingencies and transactions. While many analytical methods have been proposed as suitable for on-line DSA, because of the complex non-linear nature of power systems, only a handful of these methods have proven to be practical and have reached the stage of actual on-line implementation [6, 8]. Most power system operators have recognized the need for, and the value of, on-line DSA systems and are implementing them as add-ons to existing energy management systems (EMS) or as a requirement in the procurement of upgraded or new EMS systems. Despite the sophistication of state-of-the-art on-line DSA systems, there remains a tremendous opportunity for added functionalities and the integration of Intelligent Systems (IS) with these applications. It is anticipated that the use of IS can greatly improve the efficiency, reliability, and robustness of on-line systems. The approach is based on a PV Curve method combined with Modal Analysis [4]. Typically, a PV curve is created by increasingly stressing the system and solving a power flow at each new loading point. When the power flow fails to converge, the nose of the PV curve has been reached and this point corresponds to the stability limit for that particular imposed stress. Contingencies can also be applied at points along the PV curve to generate post-contingency.

The approach was successfully implemented and tested on the Polish Power Grid Operator - PSE Operator S.A. A database with user-friendly interface for storing/retrieving result of modal analysis for each scenario/contingency was designed. VCA identification was carried using a clustering method was developed. For the studied scenarios and contingencies in the PSE Operator S.A. system two VCAs were identified. For each of the identified VCA, a set of reactive resources, e.g. generators (RRG), was also identified for which the exhaustion of their reactive power reserve resulted in voltage collapse in the VCA.

Since speed of analysis is critical for on-line applications, this work includes also a scheme whereby VCAs can be identified using decision tree (DT) techniques from on-line system snapshot. The database can be used to construct DTs (off-line) for each of the identified VCAs using key system attributes. The relevant attributes can be extracted from a system snapshot, obtained on-line, which then can be dropped on DTs to determine which VCAs can exist in the present system condition. In addition, it is proposed to investigate the possibility of predicting the required reserve using regression trees (RT) constructed (off-line) for each of the identified VCAs using key system attributes. DTs and RTs have the added benefit of identifying the most important parameters associated with a specified outcome (such as instability). This is expected to provide valuable information to power systems operators.

1.1 Proposed Approach

Assessing and mitigating problems associated with voltage security remains a critical concern for many power system planners and operators. It is well understood that voltage security, and particular voltage stability, is driven by adequacy level of reactive power support in a system. Therefore, it is of particular interest identification

those areas in the system that may suffer reactive power deficiencies under various system operation and contingency conditions. Establishing the reactive power reserve requirements in these areas to ensure system integrity is of paramount importance. The proposed approach is based on a PV Curve method combined with Modal Analysis. The general approach is as follows:

- A system operating space is defined based on a wide range of system load conditions, dispatch conditions, and defined transactions (source-to-sink transfers).
- A large set of contingencies is defined which spans the range of credible contingencies.
- Using PV curve methods, the system is pushed through every condition, under all contingencies until the voltage instability point is found for each condition.
- To identify the VCA for each case using modal analysis: At the point of instability for each case (nose of the PV curve) modal analysis is performed to determine the critical mode of instability as defined by a set of bus participation factors corresponding to the zero eigenvalue (bifurcation point).
- The results of the modal analysis will be placed in a database for analysis using data mining methods to identify the VCAs and track them throughout the range of system changes.
- The reactive reserve requirements for selected VCA will then be established.

The network constraints are expressed in the following linearized model around the given operating point [1]:

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = \begin{bmatrix} J_{P\theta} & J_{PV} \\ J_{Q\theta} & J_{QV} \end{bmatrix} \begin{bmatrix} \Delta\theta \\ \Delta V \end{bmatrix}, \quad (1)$$

where: ΔP - incremental change in bus real power,

ΔQ - incremental change in bus reactive power,

$\Delta\theta$ - incremental change in bus voltage angle,

ΔV - incremental change in bus voltage magnitude,

$J_{P\theta}, J_{PV}, J_{Q\theta}, J_{QV}$ - are Jacobian sub-matrices.

The elements of the Jacobian matrix give the sensitivity between power flow and bus voltage changes. While it is true that both P and Q affect system voltage stability to some degree, we are primarily interested in the dominant relationship between Q and V. Therefore, at each operating point, we may keep P constant and evaluate voltage stability by considering the incremental relationship between Q and V. This is not to say that we neglect the relationship between P and V, but rather we establish a given P for the system and evaluate, using modal analysis, the Q-V relationship at that point. Based on the above consideration the incremental relationship between Q and V can be derived from Equation 1 by letting $\Delta P = 0$:

$$\Delta Q = J_R \times \Delta V \quad (2)$$

Sensitivity matrix J_R^{-1} is a full matrix whose elements reflect the propagation of voltage variation through the system following a reactive power injection in a bus.

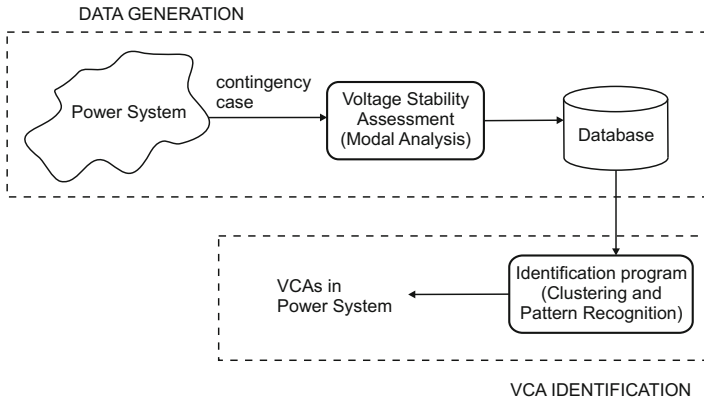


Fig. 1 VCA Identification Process

2 VSA Identification

In the presented approach, the power system is stressed to its stability limit for various system conditions under all credible contingencies. At the point of instability (nose of the PV curve) modal analysis is performed to determine the critical mode of voltage instability for which a set of bus participation factors (PF) corresponding to the zero eigenvalue (bifurcation point) is calculated. Based on these PFs, the proposed method identifies the sets of buses and generators that form the various VCAs in a given power system. It is assumed that for a given contingency case, buses with high PFs including generator terminal buses, form a VCA. This suggests that each contingency case might produce its own VCA. In practice, however, the large number of credible contingency cases generally will produce only a small number of VCAs because several contingencies are usually related to the same VCA. The proposed identification procedure applies heuristic rules to:

- group contingencies that are related to the same VCA;
- identify the specific buses and generators that form each VCA (see Fig. 1).

The following is a brief description of the proposed VCA identification program. The program processes the sets of buses and generators corresponding to the PFs obtained from the MA for each system condition and contingency case. Then contingency cases are grouped together if their sets of bus PFs are similar. To carry out this contingency clustering process, first a 'base/seed' set of VCA buses is selected. Then, all the other sets corresponding to different contingency cases are compared against this base set to determine if they are similar. Contingencies are clustered if their sets of bus PFs are similar. Finally, the program identifies the sets of buses and generators that are common to all contingencies of each cluster. Those sets of buses and generators form the VCAs of the power system.

2.1 *Heuristic Rules for Base Selection and Similarity Measurement*

2.1.1 Selection of a Base for Clustering Process

Since each contingency case is unique, a better approach to select the Set for Further Analysis (SFA) buses is to base it on the characteristics of each contingency. Generator terminal buses are PV type buses and thus are not included in the reduced Jacobian matrix. Therefore, PF cannot be calculated for a generator terminal bus until the generator exhausts its reactive reserves, which is marked as a Q -limited (QL) bus, and it becomes a PQ type bus. The number of QL buses, characteristic for each contingency, determines the selection of SFA buses. Then several contingency clusters C_k are constructed. The generator buses with the highest frequencies are selected to represent the cluster C_k reactive reserves and are denoted as GEN_k . From the VCA identification process presented in above section we can observe that clustering is carried out twice:

1. Clustering contingency cases based on SFAs,
2. Clustering C_k based on GENs.

Each clustering process starts with the selection of a base set for the cluster. Then any other set is compared to this base to evaluate whether they are similar. Both clustering processes are shown in the diagram of the VCA identification program in Fig. 2. Data Flow Diagram for VCA Identification Program.

Two different criteria for the selection of a base SFAX set were tested:

1. Largest contingency (SFA). After the SFAs are found, the number of buses in each SFA is counted. The SFA with the highest number of buses is selected as the SFAX base for a cluster and then similar SFAs are grouped together.
2. Most severe contingency (SFA). As part of the voltage stability assessment of the system, we also compute the margin for each contingency case. The SFA corresponding to the contingency with the smallest margin is selected as the base of the cluster. Then similar SFAs are grouped together.

For clustering the GEN set with the highest number of generator-buses is selected as the base GEN_x of a cluster. Then similar GENs are grouped together.

2.1.2 Measure of Similarity between Sets

Whether we are dealing with SFAs or GENs the measure of similarity is the same. First the numbers of buses in the base sets SFAX or GEN_x as well as the SFAs or GENs sets for all cases are counted. Then the elements of set- i (either SFA_i or GEN_i) are compared with the elements of the base set (either SFAX or GEN_x). The number of common elements C is counted and compared with the similarity threshold T . If the number of common elements C is greater than the threshold T , then set- i and the base set are considered being similar. The similarity threshold T

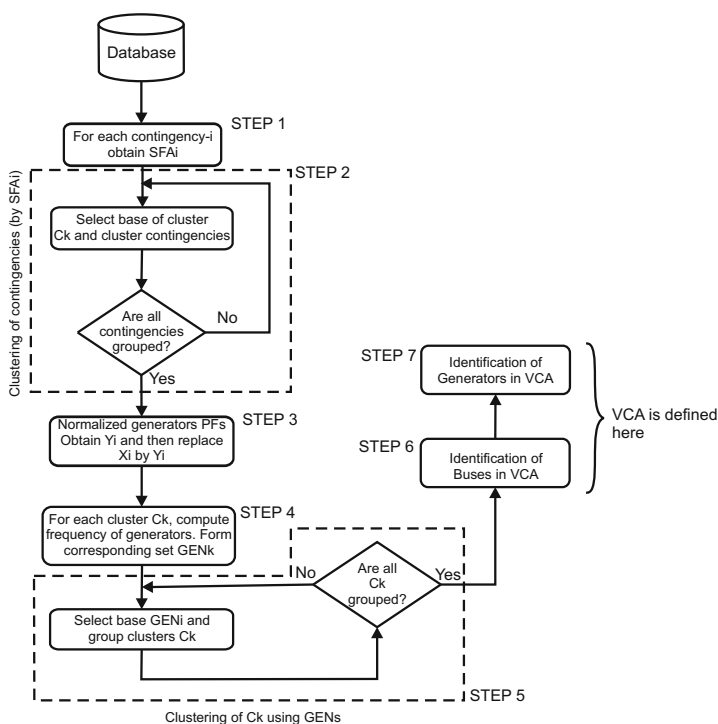


Fig. 2 Data Flow Diagram for VCA Identification Program

is set as a percentage of the number of elements of in the largest set (set- i or the base set). If all elements of the smaller set (base or set- i) are included in the larger set then those sets are considered being similar. The pseudo code for checking sets similarities is as follows:

- > Compute B =number of elements in base;
- > Compute R =number of elements in set- i ;
- > Compute maximum number of elements $M=\max(B,R)$;
- > Compute threshold for common elements $T=\phi M$;
- > Compute number of common elements between base and set- i C =common elements;
- > If $C \geq T$ then base and set- i are similar;
- > If $C < T$ then: Denote the set (base or set- i) with the lowest number of elements by S ;
- > If all elements in this smallest set are included in the largest set then sets are similar;
- > otherwise sets are not similar.

Note: The factor ϕ represents a similarity threshold. This similarity threshold is used to compute the threshold for common elements (T). The value of T depends not only of ϕ but also in the number of elements in the largest set. If the number of common elements C is equal to or greater than T , then the two sets being compared are considered to be similar. Based on computational experience $\phi = 0.50$ is used. That means that two sets are similar only if the number of common elements is equal to or greater than 50% the number of elements of the largest set.

3 Use of Decision Trees for On-Line VSA Assessment

An intelligent system framework for the application of decision trees for on-line transient stability assessment was described in [5]. A similar approach is described here for use in the assessment of voltage stability to determine VCAs and required reactive reserves that must be maintained to ensure security of each VCA. The overall architecture is shown in Fig. 3. In this Figure,

- Path D+E represents the conventional on-line VSA cycle
- Path F+G takes the conventional VSA output and computes the VCAs for all scenarios
- Path A+C combined with Path H creates a new object for the learning database including pre-contingency power flow conditions (A+C) and the corresponding VCA (H). This is a learning step used to add more information to the learning database.
- Path I represents building and rebuilding of the decision trees
- Path A+B+J represents the real-time use of the intelligent system in which a power flow from the state estimator is "dropped" on the decision trees

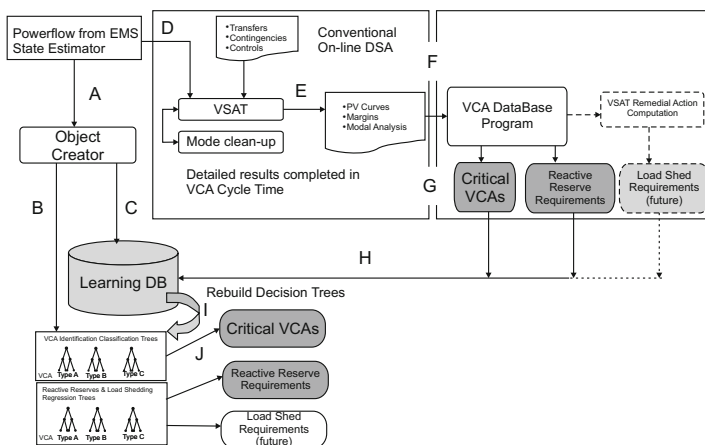


Fig. 3 Path for VCA Analysis Using Decision Trees

3.1 Building the Decision Trees for VCA Determination

The objective is to develop decision trees that can be used to determine what VCA(s) may be present for a given system condition. Generally, there are two types of decision trees [2]:

- classification trees (CT) are built for classifying data samples in terms of categorical target attribute (response variable)
- regression trees (RT) have continuous target attribute (response variable)

In the data set used for decision trees, one attribute is named *target attribute* or *goal attribute* representing decisions for classification, regression or prediction, and the others are *candidate attributes* (describing pre-decision conditions). The first step is to construct a database of the structure shown in Fig. 4.

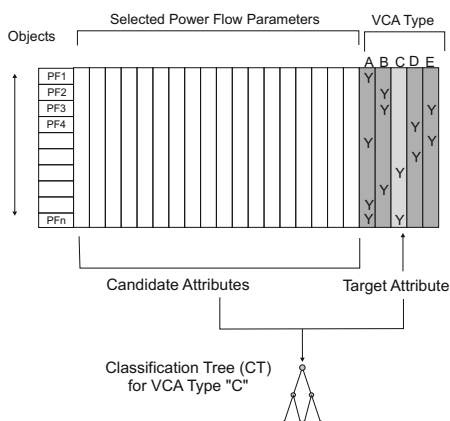


Fig. 4 Creation of Decision (Classification) Tree

Each row of the database (referred to as "objects") represents the pre-contingency conditions of a system scenario (base case condition, contingency, and transfer) for which the stability limit and VCA was found (as described in the previous section). For each scenario, the columns of the DB (referred to as "candidate attributes") contain key power base case parameters that define the system condition, and one additional column (referred to as "target attributes") for each of the different VCAs found in the full analysis space using the VCA identification process described above. The any of the VCA types occur for the given object, a "Y" (for "yes") is entered in the column, otherwise it is left blank or entered as "N" for "no". One decision tree is needed for each of the VCAs determined in the VCA grouping process. For example if VCA identification (using the VCA database) indicates five VCAs (which can be referred to as type A, B, C, D, and E, for example) five trees will be needed - one for each VCA type. When used, each DT will indicate whether a specific VCA is likely for the system condition being tested.

To build a DT for a selected VCA type, the column corresponding to the VCA type is selected as the "Target Attribute" and the other VCA columns are ignored. The DT building software is then given the entire database and a DT developed for each VCA type. A similar approach is used for developing DT for reactive reserve requirements. The required reactive reserves required for each VCA are placed in the target attribute columns and new DT computed. In the case of VCA identification, the decision trees are actually *classification trees* (with a "yes" or "no" binary output) whereas the decision trees for reactive power reserve requirements are *regression trees* (with continuous outputs which indicated the actual Mvar value required for reserves). Once the trees are developed, they are made available in the on-line system as small executable programs. Graphically, a typical decision tree for a practical power system appears as shown in Fig. 5

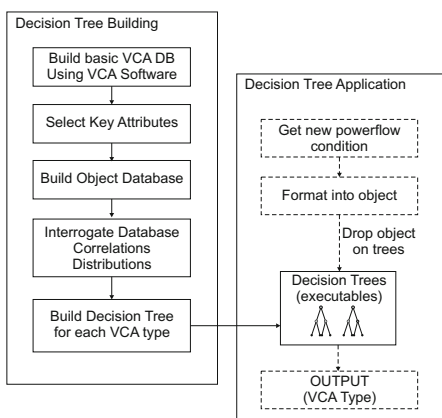


Fig. 5 Decision Tree Procedure for Determining VCA Types

When a snapshot is taken from the real-time system, the full on-line DSA engine is started as usual (Path D+E). In parallel, the snapshot is first formatted into a "new object" to contain the attributes used in the DB (Path A). Next, the DB is searched using a nearest neighbor (KNN) routine to ensure that the new object is within the scope of the DB; if it is not, the DT path is not used and conventional DSA is used. The new object is then "dropped" on all the decision trees (Path B) and the VCAs that are present are indicted (Path J), virtually instantaneously, as a "Yes" or "No" outcome of each of the decision trees. As the full DSA solution completes for all contingencies, the VCA database is grown by adding new calculation results and the VCA identification process can be updated at anytime as needed. Similarly, the IS database is appended with the new objects and outcomes, and new DTs can be built at anytime. This process is critical to ensure the system "learns" over time.

4 VCA Testing for Polish Power Grid System Operators (PSE)

The Institute of Electric Power Engineering and The Electric Power Research Institute completed the VCA testing for PSE. The result of VCA testing was encouraging and well received. PSE is to setup additional scenarios for testing and Institute of Electric Power Engineering will assist PSE in analyzing the result of VCA. The detail of the testing in Warsaw is summarized below along with the result obtained. Comprehensive testing of the software requires several steps:

- Judicious selection of practical base cases, power transfers, and contingencies to be studied.
- Tuning of the base cases to be suitable for transfer analysis.
- Setup of all required data files.
- Running VSAT to perform PV analysis and model analysis.
- Analysis of the PV results and resolution of any problems such as local modes or criteria violations.
- Running of the VCA program to generate the VCAs.
- Analysis of the results of the VCA program including VCA, critical contingencies, controlling generators, and reactive reserve requirements.

5 Conclusion

A highly automated method has been developed for the identification of areas prone to voltage instability (voltage control areas or VCAs) in practical power system models. For a wide range of system conditions and contingencies, the technique can identify the buses in each VCA and identify VCAs which are common for a set of contingencies and/or conditions. In addition, the method identifies the generators which are critical to maintaining stability for a given VCA. The method, which is based on modal analysis, has been successfully implemented in a commercial grade software tool and has been tested on a large practical system with good results.

References

- [1] Aumuller, C.A., Saha, T.K.: IEEE Trans. Power. Syst. 18, 1157–1164 (2003)
- [2] Liu, H., Bose, A., Vencatasubramanian, V.: IEEE Trans. Power. Syst. 15, 1137–1141 (2000)
- [3] Schlueter, R.A.: IEEE Trans. Power. Syst. 13, 1423–1438 (1998)
- [4] Schlueter, R.A., Liu, S., Ben-Kilian, K.: IEEE Trans. Power. Syst. 15, 1105–1111 (2000)
- [5] Tovar, G.E., Calderon, J.G., de la Torre, V.E., Nieva, F.I.: Power Systems Conference and Exposition, PSCE 2004, New York City, October 10-13 (2004)
- [6] Verma, M.K., Srivastava, S.C.: IEE Proc.-Gener. Trans. Distrib. 152, 342–350 (2005)
- [7] Voltage Stability Assessment: Concepts: Concepts, Practices and Tools. IEEE Power Engineering Society, Power System Stability Subcommittee Special Publication (August 2002)
- [8] Zhong, J., Nobile, E., Bose, A., Bhattacharya, K.: IEEE Trans. Power Syst. 19, 1555–1561 (2004)

A Review: Approach of Fuzzy Models Applications in Logistics

Dragan Simić and Svetlana Simić

Abstract. Logistics is the process of managing the flow and storage of material and information across the entire organisation with the aim to provide the best customer service in the shortest available time at the lowest cost. Long haul delivery, warehousing, fleet maintenance, distribution, inventory and order management are all examples of logistics problems. This paper outlines some current approaches of fuzzy models which are implemented in the terms of potential benefits gained in logistics domain in order to mitigate the uncertainty and risks of the current business turbulent environment and global world financial crises.

1 Introduction

Logistics is one of the vital supporting functions in manufacturing and commercial companies. Global logistics networks - supply chain management has experienced significant changes in last two decades. Conscious management of this function has a significant contribution to the success of the company. Small and medium sized companies especially, tend to undervalue the importance of logistics and do not apply appropriate logistics solutions. Thus, to stay competitive, companies rely on sophisticated information systems and logistics skills, as well as accurate and reliable forecasting systems.

The rest of the paper is organised as follows. The following section overviews the basic approach of fuzzy logic while Section 3 elaborates the logistics and supply chain management. Section 4 shows some typical implementation of fuzzy

Dragan Simić

University of Novi Sad, Faculty of Technical Sciences,

21000 Novi Sad, Serbia

e-mail: dsimic@eunet.rs

Svetlana Simić

University of Novi Sad, School of Medicine, 21000 Novi Sad, Serbia

e-mail: dsimic@eunet.rs

logic in logistics domain. Section 5 describes some future work for development and implementation of fuzzy logic and fuzzy hybrid models in logistics and supply chain management. Section 6 concludes the paper.

2 Fuzzy Logic

Fuzzy set theory was initiated by Lotfi Zadeh in 1965. In the 1970s a complete theory of evidence dealing with information from multiple sources was created [14]. It provides a mathematical framework to treat and represent uncertainty in the perception of vagueness, imprecision, partial truth, and lack of information [16]. The basic theory of fuzzy logic supplies mathematical power for the emulation of the thought and perception processes [15]. Fuzzy systems are very useful not only in situations involving highly complex systems but also in situations where an approximate solution is warranted [14]. To deal with qualitative, inexact, uncertain and complicated processes, the fuzzy logic system can be well-adopted since it exhibits a human-like thinking process [6].

Fuzzy logic is a mathematical formal multi-valued logic concept which uses fuzzy set theory. Its goal is to formalize the mechanisms of approximate reasoning. Fuzzy logic has widely been applied in various areas. Fuzzy control is one prominent example. In fuzzy control, data is characterised by linguistic variables and expert knowledge (If-Then-rules). These variables are mapped into rule bases. In fuzzy control these bases can be used for logical inferences for controlling purposes [11]. One of the reasons for the success of fuzzy logic is that the linguistic variables, values and rules enable the engineer to translate human knowledge into computer valuable representations seamlessly [16]. The treatment of data from a linguistic viewpoint is a major consideration in fuzzy set theory.

Fuzzy logic is different from probability theory because fuzzy logic is deterministic rather than probabilistic. Imprecision is modelled via fuzzy sets, linguistic variables, membership functions, inferences and de-fuzzification. These concepts are all handled in an entirely deterministic manner. There are various forms of formal fuzzy logic, fuzzy set theory and fuzzy control systems. Uncertainty in the membership functions in fuzzy set theory, i.e. uncertainty about the actual value of a membership function, has been addressed by type-2 fuzzy sets [17]. Fuzzy logic operators (such as fuzzy 'and', 'or', 'not' operators) and de-fuzzification (i.e. the transformation of a fuzzy set into a crisp value), can be modelled in various ways. A prominent example how fast growing and complex the field of fuzzy logic has become can be seen from the t-norm, which is a non-classics logic operator used for fuzzy conjunctions interpretation. There are a plethora of t-norm fuzzy logics which are discussed in [7].

From the area of data modelling, fuzzy sets have not only been extended to data summarisation by developing more abstract concepts and fuzzy gradual rules, but have also been applied to pattern recognition, as well as fuzzy clustering algorithms. In addition, a fuzzy multi-criteria decision-making algorithm has been developed for the network reconfiguration problem. It has been implemented in a proof-of-concept tool and applied to multi-criteria problems successfully [3].

3 Logistics

Supply chain management could be defined as the management of upstream and downstream relationships with suppliers and customers to deliver superior customer value at less cost to the supply chain as a whole [8]. Some other authors described the supply chain management as a plan and control of all the processes that link partners in a supply chain together in order to meet end-customers' requirements [5].

As the sub-process of supply chain management, logistics deals with planning, handling, and control of the storage of goods between manufacturer and consumer [2]. Next well-known definition of logistics best describes its contemporary meaning as the strategic management of movement, storage, and information relating to materials, parts, and finished products in supply chains, through the stages of procurement, work-in-progress and final distribution [9].

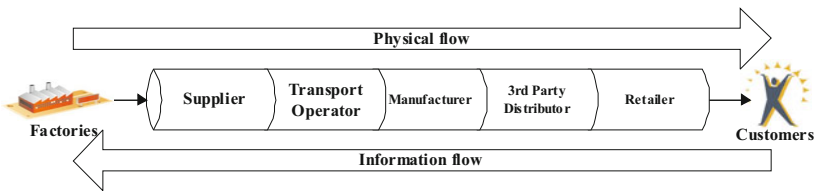


Fig. 1 The concept of supply chain

The concept of supply chain refers to the idea of developing a logistics pipeline approach for finished goods to transfer through the supply chain, and it is presented in Figure 1. The supply chain highlights the close partnership from upstream supplier, transport operator and manufacturer to the downstream 3rd party distributor and retailer. Its objective is to produce and distribute the commodity in the right quantity, to the right place, and at the right time to minimise overall cost while maintaining customer satisfaction.

4 Applications of Fuzzy Logic Models

In this section the novel applications of fuzzy logic models in logistics domain will be described. Fuzzy logic models have been researched and implemented in logistics for a very long time, but, in this paper, some of the cutting edge researches will be presented. Most of these researches and implementations were conducted in year 2010. and published in year 2011. In the following subsections they will be presented in different logistic branches such as: 1) fuzzy application of an expert system for some logistic problems [12]; 2) a multi-criteria decision making approach for location planning for urban distribution centers under uncertainty [1]; 3) very

large street routing problem with mixed transportation mode - fuzzy approach [10]; 4) design and implementation of a fuzzy expert decision system for vendor selection [13]; 5) fuzzy decision making approach to IT outsourcing supplier selection [4].

4.1 Fuzzy Application of an Expert System for Some Logistic Problems

In this study [12], the instance of a company planning transport contracts with own trucks is analysed, but when needed, the firm may also lease some external transportation means. It may aid the managers with their financial and technical aspects analysis of transport contracts. Technical aspects in the voyage estimation include such variables as: the distances between the company warehouses and customers, the cargo capacity of the transport means, speed, place distances, fuel consumption, the load patterns, etc. The listed technical details have rather a stable, invariant character. The financial aspects include costs of the transportation means and time, freight rates for different types of cargo and other expenses. Financial factor, which includes the costs of fuel and labor, may vary because of some dependencies on weather conditions, and the time spent waiting for unloading/loading, etc.

Because of the fact that some of the transportation contracts' characteristics are imprecise or unpredictable, the systems of fuzzy logic and probability methods are suggested. They proposed fuzzy expert system enabling the choice of the appropriate customer order to maximise the company's profit.

Decision support system will create a suggestion on how to select a client for cargo transportation between a company warehouse and a client. The system will perform deduction after analyzing following voyage features:

1. Transportation cost defined as aggregation of a distance between a customer and a warehouse, fuel consumption of a transportation mean and possible obstacles expected during a planned voyage.
2. Order capacity value defined as aggregation of order capacity declared on paper and a kind of client regarding its importance for the company.
3. Warehouse bulk resources.
4. Warehouse owner profit defined as aggregation of a client's profile and transportation cost.

The most of the analysed features are defined with description based on fuzzy sets. A few of the above mentioned features are continuous variables, while the probability functions represent discrete variables. The membership functions used for the features described are shown: a) the distance between a customer and a warehouse, order capacity, truck fuel consumption, transport cost; b) the client profile variable has been described also by means of fuzzy sets; c) two fuzzy sets (empty and full) were used to describe the embarkation rate of a truck; d) the warehouse owner profit and transportation conditions have also been described.

4.2 Fuzzy Approach for Location Planning for Urban Distribution Centers

Location planning for urban distribution centers is vital in saving distribution costs and minimizing traffic congestion arising from goods movement in urban areas [1]. In recent years, transport activity has grown tremendously and this has undoubtedly affected the travel and living conditions in urban areas. Considering this growth in the number of urban freight movements and their negative impacts on city residents and the environment, municipal administrations are implementing sustainable freight regulations like restricted delivery timing, dedicated delivery zones, congestion charging etc. With the implementation of these regulations, the logistics operators are facing new challenges in location planning for distribution centers. If distribution centers are located close to customer locations, then it will increase traffic congestion in the urban areas. If they are located far from customer locations, then the distribution costs for the operators will be very high. Under these circumstances, the location planning for distribution centers in urban areas is a complex decision that involves consideration of multiple criteria like: maximum customer coverage, minimum distribution costs, least impacts on city residents and the environment, and conformance to freight regulations of the city.

The proposed approach is a multi-criteria decision making for location planning for urban distribution centers under uncertainty. It involves identification of potential locations, selection of evaluation criteria, use of fuzzy theory for quantifying criteria values under uncertainty and select the best location for implementing an urban distribution center. Sensitivity analysis is performed to determine the influence of criteria weights on location planning decisions for urban distribution centers. It is essential to represent criteria for location selection: 1) accessibility - access by public and private transport modes to the location; 2) security - security of the location from accidents, theft and vandalism; 3) connectivity to multimodal transport - with other modes of transport highways, railways, seaport, airport; 4) costs - in acquiring land, vehicle resources, drivers, and taxes for the location; 5) environmental impact - impact of location on the environment (air, pollution, noise); 6) proximity to customers - distance of location to customer locations; 7) proximity to suppliers - distance of location to supplier locations; 8) resource availability - availability of raw material and labor resources in the location; 9) conformance to sustainable freight regulations - ability to conform to sustainable freight regulations imposed by municipal administrations (restricted delivery, hours, special delivery zones); possibility of expansion - ability to increase size to accommodate growing demands; 11) quality of service - ability to assure timely and reliable service to clients.

In such system, these criteria are used to model systems that are difficult to define precisely; parameters are defined using linguistic terms instead of exact numerical values. Then, conversion scale is applied to transform the linguistic scale into fuzzy number.

4.3 *Very Large Street Routing Problem with Mixed Transportation Mode*

In recent years many centralisation activities in postal delivery companies to improve performance of their processes can be noticed. It is a reaction to globalization and current financial problem. Part of this process is lowering of number of Distribution Centres (DC) and creating larger centres. Postal delivery in agglomerations with large number of customers is one of the most difficult operations problems faced by local delivery companies [10].

To continue this discussion about providers of logistics service, it is possible to conclude, that they are under a big pressure to lower their expenses. One way to accomplish this task is centralisation of logistics activities. This creates distribution centers with a large number of customers.

This problem is often called a Street Routing Problem (SRP). In this research a survey of aggregation heuristics that can be used for a solution of Very Large SRP (VLSRP) is used. Performance of heuristics has been evaluated based on real data. This paper presents several approximations of length for a SRP with mixed transportation mode and compares them with published approximations used for Vehicle Routing Problem (VRP) or Traveling Salesman Problems (TSP).

In today's postal delivery there is a trend to centralize operations of distribution to small number of centers. This creates a situation in which many customers are served from one central depot. Centralizing all delivery operations to one depot with many customers has its advantages and disadvantages. One key advantage to centralization is the ability to use more advanced and expensive technologies for preparation of deliveries, before they are picked up. The delivery preparation usually takes more than 10% of the total working time of mail carriers. By shortening this time, combined with good transportation methods, one can allow more time for a delivery and/or reduce number of mail carriers. On the other hand, the process of restructuring requires some large investments in automatic sorting lines for every DC. Other expenses are related to physical distribution of mail carriers. Currently they use a public transport or foot to get to their districts. In new scenario there is a plan to use cars for moving the mail carriers to and from districts.

Now, part of Fuzzy Cluster Aggregation (FCA) will be described in details:

1. Estimate the minimum number of clusters p needed for serving all the customers.
2. Locate p medians, so they are uniformly distributed in the serviced area.
3. Create p clusters of customers around these medians using clustering by "fuzzy c-means". The membership of customers in each cluster is set as a triangle fuzzy number, which is based on route distance between the cluster median and customer.
4. For each cluster, approximate the route length and estimate the average service time.
5. If there are many clusters (more than 20%) which have their service time over, increase the number of clusters p and go to step 2, otherwise end the algorithm with the resulting clusters and use them as SRP districts.

The solution that FCA gives can be infeasible. To make it feasible one can lower the threshold, or make and exchange customers between the aggregates to make the solution feasible.

4.4 Implementation a Fuzzy Expert Decision Support System for Vendor Selection

The explosive growth in “business to business” commerce is expected to revolutionise the transaction process between buyers and sellers. Effective purchasing and supply management can contribute significantly to the success of most companies. Procurement and supply management are one of the most significant parts in Engineering, Procurement, and Construction (EPC) contracts [13].

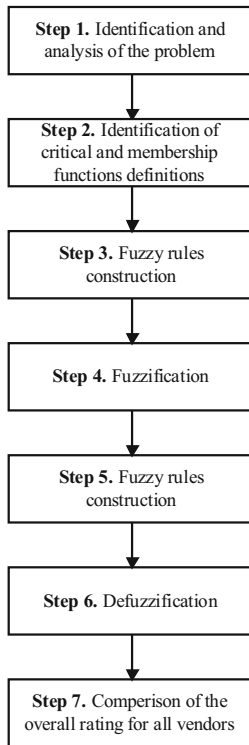


Fig. 2 Fuzzy inference process for vendor selection

When a supplier selection decision needs to be made, the buyer generally establishes a set of evaluation criteria that can be used to compare potential sources. The basic criteria typically utilized for this purpose are pricing structure, delivery (lead-time and reliability), product quality, and service (personnel, facilities, research and

development, capability). Frequently, these evaluation criteria conflict with one another. In addition, the importance of each criterion varies from one purchase to the next. This situation can be more complicated further by the fact that some of the criteria are primarily quantitative (price, quality, etc.) and some are qualitative (service). The overview of the framework is shown in Figure 2. There are 7 fundamental steps in the development of a fuzzy expert decision support system. For using the same system for all bids in all projects a weigh for each of the factors price is defined, quality and delivery time and the final score that is a number between 0 and 1 will be the average of the system outputs considering all the factors.

4.5 *Fuzzy Decision Making Approach to IT Outsourcing Supplier Selection*

Industrial organizations are constantly in search for new solutions and strategies to develop and increase their competitive advantage. Outsourcing is one of these strategies that can lead to greater competitiveness. Briefly, it can be defined as a managed process of transferring activities to be performed by others. Information technology (IT) outsourcing means that the physical and/or human resources related to an organization's ITs are supplied and/or administered by an external specialized provider [4]. IT outsourcing is often more efficient than developing systems internally because production costs are lower with outsourcing. The provider obtains scale economies from mass-producing its services and distributing its fixed costs among a great number of end-user clients. Outsourcing IT can include data centers, wide area networks, applications development and maintenance functions, end-user computing and business processing.

Table 1 Main steps of the proposed methodology [13]

S	1	Determining evaluation criteria
	2	Determining the alternatives
T	3	Determining design and system ranges (FRs)
	4	Translating linguistic terms into fuzzy numbers)
E	5	Aggregating expert opinions
	6	Determining criteria weights
P	7	Calculating information contents
	8	Calculating weighted information contents
S	9	Ranking the alternatives

The purpose of IT outsourcing supplier selection is to determine the optimal service provider that offers the best all-around package of products and services for the customer. A set of evaluation criteria has to be defined prior to evaluation of prospective IT suppliers that have compatible goals, appropriate skills, effective motivation,

and complementary strategic orientations. This is because managers must find ways to develop win-win deals (both partners benefit) for this outsourcing to be successful. Thus, it is critical for managers to identify and understand effective supplier selection criteria prior to entering into strategic outsourcing. Based on in-depth study of IT supplier selection criteria six main evaluation criteria are determined and used: 1) Technological capability; 2) Profitability of supplier; 3) Relationship closeness; 4) Total cost; 5) Service quality; 6) Reputation of supplier.

5 Conclusion and Future Work

This paper outlines some current approaches of fuzzy models which are implemented in the terms of potential benefits gained in logistics domain in order to mitigate the uncertainty and risks of the current business turbulent environment and global world financial crises.

Therefore, first a short overview of some logistics problems which can be discussed in every modern company is presented. Then, a multi-criteria decision making approach for location planning for urban distribution centers under uncertainty and current business turbulent environment is presented. Also, urban distribution centers are presented, which are closely related to - large street routing problem with mixed transportation mode - which is presented in fuzzy approach. It is also necessary, if logistics problems are considered, to implement fuzzy based decision support system, for vendor selection. It has lately been very common for companies to concentrate their business around their core activities. Very common practice is outsourcing of transportation and warehousing but lately corporate IT functions are outsourced as well. Also, researches on implementation of different artificial intelligence techniques can be applied to almost all logistic activities of the supply chain management.

References

- [1] Awasthi, A., Chauhan, S.S., Goyal, S.K.: *Mathematical and Computer Modelling* 53, 93–109 (2011)
- [2] Back, T., Hammel, U., Schwefel, H.P.: *IEEE Transactions on Evolutionary Computation* 1, 3–17 (1997)
- [3] Bosc, P., Kraft, D., Petry, F.: *Fuzzy Sets and Systems* 156, 418–426 (2005)
- [4] Buyüközkan, G., Ersoy, M.: *Engineering and Technology* 55, 411–415 (2009)
- [5] Christopher, M.: *Logistics and Supply Chain Management*, 2nd edn. Prentice Hall, Norfolk (2004)
- [6] Du, T.C., Wolfe, P.M.: *Computers in Industry* 32, 261–272 (1997)
- [7] Esteva, F., Godo, L., Montagna, F.: *Archive for Mathematical Logic* 40, 39–67 (2001)
- [8] Goldberg, D.E.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Massachusetts (1989)
- [9] Harrison, A., Hoek, R.: *Logistics Management and Strategy*, 2nd edn. Prentice Hall, Essex (2005)

- [10] Matis, P., Kohni, M.: *Central European Journal of Operations research* (2010), doi: 10.1007/s10100-010-0159-2
- [11] Pieczyński, A., Robak, S.: Application of an Expert System for Some Logistic Problems. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) *ICAISC 2010. LNCS (LNAI)*, vol. 6114, pp. 630–637. Springer, Heidelberg (2010)
- [12] Petry, F.E., Zhao, L.: *Fuzzy Sets and Systems* 160, 2206–2223 (2009)
- [13] Ramezani, M., Montazer, G.A.: Design and implementation of a fuzzy expert decision support system for vendor selection. In: Manolopoulos, Y., Filipe, J., Constantinopoulos, P., Cordeiro, J. (eds.) *ICEIS 2006 - Proceedings of the Eighth International Conference on Enterprise Information Systems: Databases and Information Systems Integration* (2006)
- [14] Ross, T.J.: *Fuzzy Logic with Engineering Applications*, 2nd edn. John Wiley and Sons, West Sussex (2004)
- [15] Sinha, N.K., Gupta, M.M.: Introduction to soft computing and intelligent control systems. In: Sinha, N.K., Gupta, M.M. (eds.) *Soft Computing and Intelligent Systems: Theory and Applications*. Academic Press, San Diego (2000)
- [16] Tettamanzi, A., Tomassini, M.: *Soft computing: Integrating Evolutionary, Neural and Fuzzy Systems*. Springer, Heidelberg (2001)
- [17] Zadeh, L.A.: *IEEE Computational Intelligence Magazine* 3, 11–22 (2008)

Expansion of Matching Pursuit Methodology for Anomaly Detection in Computer Networks

Łukasz Saganowski, Tomasz Andrysiak, Michał Choraś, and Rafał Renk

Abstract. In this paper we present further expansion of our matching pursuit methodology for anomaly detection in computer networks. In our previous work we proposed new signal based algorithm for intrusion detection systems based on anomaly detection approach on the basis of the Matching Pursuit algorithm. Hereby, we present further modifications of our methodology and we report improved results on the benchmark data sets.

1 Problem Formulation

Anomaly detection approach is a new, emerging trend for network security especially for high-security networks (such as military or critical infrastructure monitoring networks). Such networks are currently exposed to many threats due to the fact that barriers between trusted and un-trusted network components do not successfully protect critical parts of the cyber domain. Most IDS/IPS (Intrusion Detection/Prevention Systems) cannot cope with new sophisticated malware (viruses, SQL injections, Trojans, spyware and backdoors) and 0-day attacks. Most current IDS/IPS systems have problems in recognizing new attacks (0-day exploits) since they are based on the signature-based approach. In such mode, when system does not have an attack signature in database, the attack is not detected. Another drawback of current IDS systems is that the used parameters and features do not contain all the necessary information about traffic and events in the network [1].

Łukasz Saganowski · Tomasz Andrysiak · Michał Choraś
Institute of Telecommunications, University of Technology & Life Sciences in Bydgoszcz
e-mail: luksag@utp.edu.pl

Michał Choraś · Rafał Renk
ITTI Ltd., Poznań
e-mail: mchoras@itti.com.pl

Rafał Renk
Adam Mickiewicz University, Poznań
e-mail: renk@amu.edu.pl

Therefore in our work, we present a new solution for ADS system based on signal processing algorithm. ADS analyzes traffic from internet connection in certain point of a computer network. The proposed ADS system uses the expansion of the previously developed redundant signal decomposition method based on Matching Pursuit algorithm [2][3].

The paper is structured as follows: in Section 2 the Anomaly Detection System based on adaptive approximation of signals is described in details. Then, in Section 3, the experimental scenario and results are presented and discussed. Conclusions are given thereafter.

2 Anomaly Detection System Based on Adaptive Approximation of Signals

Given overcomplete set of functions called dictionary $D = \{g_{\gamma_0}, g_{\gamma_1}, \dots, g_{\gamma_{n-1}}\}$ such that norm $\|g_{\gamma_i}\| = 1$, we can define an optimal M - approximation as an expansion, minimizing the error δ of an approximation of signal $f(t)$ by M waveforms g_{γ_i} called atoms:

$$\delta = \left\| f(t) - \sum_{i=0}^{M-1} \alpha_i g_{\gamma_i} \right\| \quad (1)$$

where $\{\gamma_i\}_{i=1,2,\dots,M-1}$ represents the indices of the chosen functions g_{γ_i} [9]. Finding such an optimal approximation is an NP -hard problem [17]. A suboptimal expansion can be found by means of an iterative procedure, such as the matching pursuit algorithm.

2.1 Matching Pursuit Overview

Matching pursuit is a recursive, adaptive algorithm for signal decomposition [9]. The matching pursuit decomposes any signal into linear expansion of waveforms which are taken from an overcomplete dictionary D . Signal f can be written as the weighted sum of dictionary elements:

$$f = \sum_{i=0}^{N-1} \alpha_i g_{\gamma_i} + R^n f \quad (2)$$

where $R^n f$ is residual in an n - term sum. In the first step of Matching Pursuit algorithm, the waveform g_{γ_0} which best matches the signal f is chosen. The first residual is equal to the entire signal $R^0 f = f$. In each of the consecutive steps, the waveform g_{γ_n} is matched to the signal $R^n f$ which is the residual left after subtracting results of previous iterations:

$$R^n f = R^{n-1} f - \alpha_n g_n \quad (3)$$

where

$$\alpha_n = \langle R^{n-1} f, g_{\gamma_n} \rangle \quad (4)$$

and

$$g_{\gamma_n} = \arg \max_{g_{\gamma_n} \in D} |\langle R^{n-1} f, g_{\gamma_n} \rangle|. \quad (5)$$

When $R^n f$ is minimized for a given $g_{\gamma_{n-1}}$, the projection between the previous residue and actual atom $\langle R^{n-1} f, g_{\gamma_{n-1}} \rangle$ is maximized. Iteratively, we obtain for N atom:

$$R^N f = f - \sum_{n=0}^{N-1} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} \quad (6)$$

where $R^n f \rightarrow 0$ when $N \rightarrow \infty$ [17]. This describes the decomposition process.

2.2 Dictionary of Gabor Functions

In the described ADS solution, we proposed a waveform from a time-frequency dictionary that can be expressed as translation (u), dilation (s) and modulation (ω) of a window function $g(t) \in L^2(\mathbb{R})$

$$g_{\gamma}(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) e^{i\omega t} \quad (7)$$

where

$$g(t) = \frac{1}{\sqrt{s}} e^{-\pi t^2} \quad (8)$$

Optimal time-frequency resolution is obtained for Gaussian window $g(t)$, which for the analysis of real valued discrete signals gives a dictionary of Gabor functions

$$g_{\gamma}(x) = C(\gamma, \varphi) g\left(\frac{x-u}{s}\right) \sin\left(2\pi \frac{\omega}{N}(x-u) + \varphi\right) \quad (9)$$

where K is the size of the signal for which the dictionary is constructed, $C(\gamma, \varphi)$ is normalizing constant used achieve atom unit energy $\|g_{\gamma}\| = 1$ and $\gamma = \{s, u, \omega, \varphi\}$ denotes parameters of the dictionary's functions. We implemented the dictionary originally proposed by Mallat and Zhang in [9], the parameters of the atoms are chosen from dyadic sequences of integers. Scale s , which corresponds to an atom's width in time, is derived from the dyadic sequence $s = 2^j, 0 < j < L$ (signal size $K = 2^L$ and j is octave). Parameters u and ω , which correspond to an atom's position in time and frequency, are sampled for each octave j with interval $s = 2^j$. In order to create overcomplete set of Gabor functions dictionary D was built by varying subsequence atom parameters: scale (s), translation (u), modulation (ω) and phase (φ). Base functions dictionary D was created with using 10 different scales and 50 different modulations.

2.3 Search in Dictionary Atoms

In the basic Matching Pursuit algorithm atoms are selected in every step from entire dictionary which has flat structure. In this case algorithm causes significant processor burden. In our ADS system dictionary with internal structure was used.

Dictionary is built from:

- Atoms,
- Centered atoms,

Centered atoms groups such atoms from D that are as more correlated as possible to each other. To calculate measure of correlation between atoms function $o(a, b)$ can be used [10].

$$o(a, b) = \sqrt{1 - \left(\frac{|\langle a, b \rangle|}{\|a\|_2 \|b\|_2} \right)^2} \quad (10)$$

The quality of centered atom can be estimated according to [11]:

$$O_{k,l} = \frac{1}{|LP_{k,l}|} \sum_{i \in LP_{k,l}} o(A_{c(i)}, W_{c(k,l)}) \quad (11)$$

$LP_{k,l}$ is a list of atoms grouped by centered atom. $O_{k,l}$ is mean of local distances from centered atom $W_{c(k,l)}$ to the atoms $A_{c(i)}$ which are strongly correlated with $A_{c(i)}$.

Centroid $W_{c(k,l)}$ represents atoms $A_{c(i)}$ which belongs to the set $i \in LP_{k,l}$. List of atoms $LP_{k,l}$ should be selected according to the Equation [12]:

$$\max_{i \in LP_{k,l}} o(A_{c(i)}, W_{c(k,l)}) \leq \min_{t \in D \setminus LP_{k,l}} o(A_{c(t)}, W_{c(k,l)}) \quad (12)$$

3 Experimental Results

Performance of our approach was evaluated with the use of four trace bases [13, 14, 15]. The test data contains attacks that fall into four main categories [11] such as:

1. DOS/DDOS: denial-of-service, e.g. syn flood,
2. R2L: unauthorized access from a remote machine, e.g. guessing password,
3. U2R: unauthorized access to local superuser (root) privileges, e.g., various "buffer overflow" attacks,
4. PROBING: surveillance and other probing, e.g., port scanning.

For experiments we chose 20 minutes analysis window because most of attacks (about 85%) ends within this time period [12]. We extracted 15 traffic features (see Table 8) in order to create 1D signals for Matching Pursuit - Mean Projection analysis. Graphic representation examples of ICMP and TCP flow/minute parameter for "normal" and "attacked" traffic is presented in Figures 112 and Figures 314.

For anomaly detection classification we used two parameters. First parameter is Matching Pursuit Mean Projection (MP-MP), second energy parameter $E_{(k)}$.

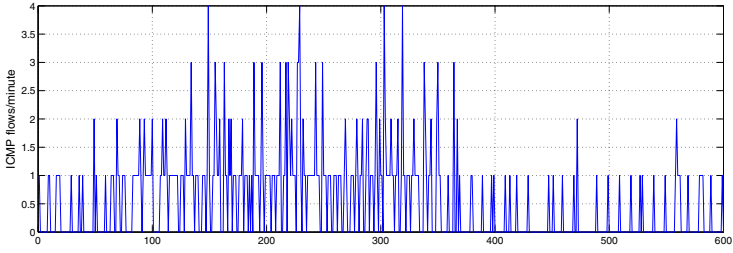


Fig. 1 ICMP flows per minute without attack for DARPA [13] Week 5 Day 1 trace

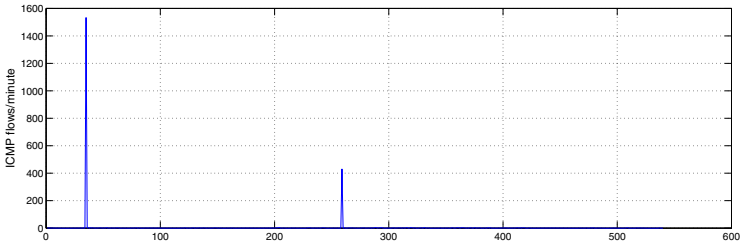


Fig. 2 ICMP flows per minute with attack for DARPA [13] Week 5 Day 1 trace

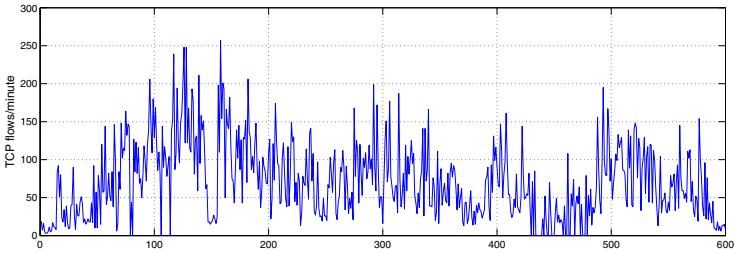


Fig. 3 TCP flows per minute without attack for DARPA [13] Week 5 Day 1 trace

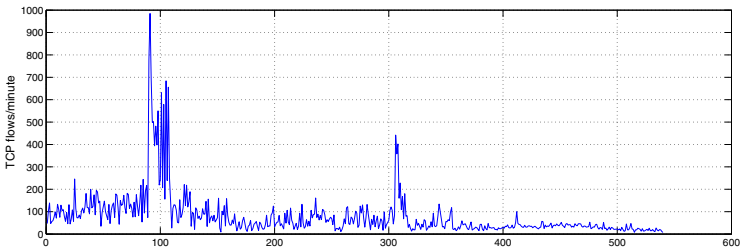


Fig. 4 TCP flows per minute with attack for DARPA [13] Week 5 Day 1 trace

Table 1 Detection Rate for W5D5 (Fifth Week, Day 5) DARPA [13] trace

Network Traffic Feature	DR[%] for MPMP	DR[%] for $E_{(k)}$
icmp flows/min.	70,58	88,23
icmp in bytes/min.	41,17	76,47
icmp in frames/min.	66,17	80,88
icmp out bytes/min.	39,70	77,94
icmp out frames/min.	66,17	77,94
tcp flows/min.	51,47	55,88
tcp in bytes/min.	69,11	92,64
tcp in frames/min.	52,94	89,70
tcp out bytes/min.	41,17	94,11
tcp out frames/min.	47,05	79,41
udp flows/min.	88,23	97,05
udp in bytes/min.	100,00	100,00
udp in frames/min.	92,64	98,52
udp out bytes/min.	92,64	94,11
udp out frames/min.	98,52	94,11

Table 2 Matching Pursuit Mean Projection - MP-MP for TCP trace (20 min. analysis window)

TCP trace (packet/second) MAWI [14]	Window1 MPMP	Window2 MPMP	Window3 MPMP	MPMP for trace	MPMP for normal trace
Mawi 2004.03.06 tcp	210.34	172.58	239.41	245.01	240.00
Mawi 2004.03.13 tcp	280.01	214.01	215.46	236.33	240.00
Mawi 20.03.2004 tcp (attacked: Witty)	322.56	365.24	351.66	346.48	240.00
Mawi 25.03.2004 tcp (attacked: Slammer)	329.17	485.34	385.50	400.00	240.00

The matching pursuits algorithm produces three important elements of information: the set of projection coefficients $\alpha = \{\alpha_0, \alpha_1, \dots, \alpha_{n-1}\}$, the set of residues $Rf = \{R^0f, R^1f, \dots, R^{n-1}f\}$ and the list of dictionary elements chosen to approximate of $f(x)$, represented as $g_\gamma = \{g_{\gamma_0}, g_{\gamma_1}, \dots, g_{\gamma_{n-1}}\}$. This three factors α , Rf and g_γ completely define the discrete signal $f(x)$. Their energies can be written according to the equation [13]:

$$E_{(k)} = \|\alpha^{(k)}\|^2 + \|Rf^{(k)}\|^2 + \|g_\gamma^{(k)}\|^2 \quad (13)$$

Table 3 Matching Pursuit Mean Projection - MP-MP for UDP trace (20 min. analysis window)

UDP trace (packet/second) MAWI [14]	Window1 MPMP	Window2 MPMP	Window3 MPMP	MPMP for trace	MPMP for normal trace
Mawi 2004.03.06 udp	16.06	13.80	17.11	15.65	16.94
Mawi 2004.03.13 udp	20.28	17.04	17.40	18.24	16.94
Mawi 20.03.2004 udp (attacked: Witty)	38.12	75.43	61.78	58.44	16.94
Mawi 25.03.2004 udp (attacked: Slammer)	56.13	51.75	38.93	48.93	16.94

Table 4 Matching Pursuit Energy parameter - for TCP trace (20 min. analysis window)

TCP trace (packet/second) MAWI [14]	Window1 $E_{(k)}$	Window2 $E_{(k)}$	Window3 $E_{(k)}$	$E_{(k)}$ for trace	$E_{(k)}$ for normal trace
Mawi 2004.03.06 tcp	7.26e+7	7.92e+7	6.36e+7	7.18e+7	7.83e+7
Mawi 2004.03.13 tcp	8.69e+7	7.80e+7	8.98e+7	8.49e+7	7.83e+7
Mawi 20.03.2004 tcp (attacked: Witty)	2.03e+8	1.46e+8	1.83e+8	1.77e+8	7.83e+7
Mawi 25.03.2004 tcp (attacked: Slammer)	3.43e+8	2.39e+8	3.51e+8	3.11e+8	7.83e+7

Table 5 Matching Pursuit Energy parameter for UDP trace (20 min. analysis window)

UDP trace (packet/second) MAWI [14]	Window1 $E_{(k)}$	Window2 $E_{(k)}$	Window3 $E_{(k)}$	$E_{(k)}$ for trace	$E_{(k)}$ for normal trace
Mawi 2004.03.06 udp	4.13e+5	2.26e+5	2.49e+5	2.96e+5	3.79e+5
Mawi 2004.03.13 udp	5.08e+5	4.61e+5	4.19e+5	4.62e+5	3.79e+5
Mawi 20.03.2004 udp (attacked: Witty)	5.19e+6	3.96e+6	5.49e+6	4.88e+6	3.79e+5
Mawi 25.03.2004 udp (attacked: Slammer)	1.76e+6	2.06e+6	1.73e+6	1.85e+6	3.79e+5

We calculate difference energy $Dif f_E$ between attacked and corresponding normal traces stored in a references database. If the value $Dif f_E$ is larger than a certain threshold t our application signalizes the attack/anomaly.

Table 6 Matching Pursuit Mean Projection - MP-MP for TCP trace with DDoS attacks (20 min. analysis window)

TCP trace (packet/second) CAIDA [15]	Window1 MPMP	Window2 MPMP	Window3 MPMP	MPMP for trace	MPMP for normal trace
Backscatter 2008.11.15	147.64	411.78	356.65	305.35	153.66
Backscatter 2008.08.20	208.40	161.28	153.47	147.38	153.66

Table 7 Matching Pursuit Energy parameter for TCP trace with DDoS attacks (20 min. analysis window)

TCP trace (packet/second) CAIDA [15]	Window1 $E_{(k)}$	Window2 $E_{(k)}$	Window3 $E_{(k)}$	$E_{(k)}$ for trace	$E_{(k)}$ for normal trace
Backscatter 2008.11.15	4.12e+7	3.41e+8	2.16e+8	1.99e+8	4.68e+7
Backscatter 2008.08.20	9.45e+7	4.17e+7	5.76e+7	6.46e+7	4.68e+7

Table 8 Proposed MP based ADS in comparison to DWT based ADS [16]. Both solutions were tested with the use of DARPA [13] testbed (results in table are for Week5 Day1 testday; DR-Detection Rate [%], FP-False Positive [%]) for MP-MP and $E_{(k)}$ energy parameter.

Traffic Feature	MP-MP DR[%]	MP-MP FP[%]	$E_{(k)}$ DR[%]	$E_{(k)}$ FP[%]	DWT DR[%]	DWT FP[%]
ICMP flows/minute	68.49	20.54	90.41	38.35	14.00	79.33
ICMP in bytes/minute	79.45	27.39	94.52	38.35	83.33	416.00
ICMP out bytes/minute	73.97	32.87	94.52	38.72	83.33	416.00
ICMP in frames/minute	78.08	27.39	84.93	34.24	32.00	112.00
ICMP out frames/minute	72.60	30.13	94.52	35.61	32.00	112.00
TCP flows/minute	89.04	34.24	98.63	39.72	26.67	74.67
TCP in bytes/minute	47.94	32.87	93.15	41.00	8.67	23.33
TCP out bytes/minute	80.82	27.39	93.15	34.24	8.67	23.33
TCP in frames/minute	36.98	26.02	95.89	36.09	2.00	36.00
TCP out frames/minute	38.35	27.39	95.89	34.24	2.00	36.00
UDP flows/minute	89.04	41.09	90.41	39.70	10.00	74.67
UDP in bytes/minute	98.63	41.09	98.63	45.02	11.33	66.67
UDP out bytes/minute	100.00	46.57	100	46.57	11.33	66.67
UDP in frames/minute	98.63	39.72	98.63	45.20	12.67	66.67
UDP out frames/minute	100.00	46.57	100	45.20	12.67	66.67

$$Diff_E(NRM, ATT) = \frac{|E_{(NRM)} - E_{(ATT)}|}{\max(E_{(NRM)}, E_{(ATT)})} \quad (14)$$

where E_{NRM} is energy factor of the normal trace and E_{ATT} is energy factor of the attacked trace.

All test bases were calculated with usage of this two parameters. $E_{(k)}$ parameter gives us better detection rates than MP-MP parameter for almost all traffic features at the the cost of a little bit higher false positive. Traffic features were calculated with the use of 1 minute time period (Table 1 - Table 8). Traffic feature called *flow* requires more explanation since there are various definitions of this metric. A flow consist of a group of packets going from a specific source to a specific destination over a time period [16]. Flows are always considered as sessions between users and services. For every network traffic feature we created "normal profile" based on DARPA training base without attacks. Results achieved for testing base were compared to normal profiles in order to detect anomalies. We compared our system to other signal processing ADS system based on Wavelet transform [16].

For the same DARPA test base we achieved better anomaly detection rate. Our system is more computational complex than system based on wavelets transform but it is possible to realize it in real time.

In Table 1 and Table 8 detection rates achieved for DARPA benchmark trace base are presented. These are results achieved for two test days. Detection results were compared to the list of attacks which should exist in this two testing days.

In Tables 2, 3 and Tables 4, 5 there are results for MAWI test base. Bold numbers in tables point to existence of anomalies/attacks in certain window. In Tables 6, 7 there are results achieved for CAIDA test base. Traces consist of DDoS attacks and every trace represents 1 hour of the network traffic.

4 Conclusions

In this paper we presented further expansion of our previously proposed matching pursuit methodology for anomaly detection in computer networks [2][3]. As proved in the performed experiments, we improved the efficiency of our system in terms of both Detection Rate and False Positives.

Our anomaly detection methodology has been designed for protecting critical infrastructures (e.g. SCADA networks monitoring energy grids), military networks and public administration networks.

Acknowledgement. This work was partially supported by Polish Ministry of Science and Higher Education funds allocated for the years 2010-2012 (Research Project number OR00012511).

References

- [1] Coppolino, L., D'Antonio, S., Esposito, M., Romano, L.: Exploiting diversity and correlation to improve the performance of intrusion detection systems. In: Proc of IFIP/IEEE International Conference on Network and Service (2009)
- [2] Saganowski, Ł., Choraś, M., Renk, R., Hołubowicz, W.: A Novel Signal-Based Approach to Anomaly Detection in IDS Systems. In: Kolehmainen, M., Toivanen, P., Beliczynski, B. (eds.) ICANNGA 2009. LNCS, vol. 5495, pp. 527–536. Springer, Heidelberg (2009)
- [3] Choraś, M., Saganowski, Ł., Renk, R., Hołubowicz, W.: Statistical and signal-based network traffic recognition for anomaly detection. *Expert Systems: The Journal of Knowledge Engineering* (2011)
- [4] Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations. *Journal of Constructive Approximations* 13, 57–98 (1987)
- [5] Gilbert, A., Muthukrishnam, S., Strauss, M.J.: Approximation of functions over redundant dictionaries using coherence. In: 14th ACM-SIAM Symposium on Discrete Algorithms (2003)
- [6] Gabor, D.: Theory of communication. *Journals Electrical Engineers* 93, 429–457 (1946)
- [7] Goodwin, M.: *Adaptive Signal Models: Theory, Algorithms, and Audio Algorithms*. Kluwer, Boston (1998)
- [8] Natarajan, B.K.: Sparse approximate solutions to linear systems. *SIAM Journal of Computation* 24, 227–234 (1995)
- [9] Zhang, M.S.: Matching Pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41(12), 3397–3415 (1993)
- [10] Jost, P., Vandergheynst, P., Frossard, P.: *Tree-Based Pursuit: Algorithm and Properties*. Swiss Federal Institute of Technology Lausanne (EPFL), Signal Processing Institute Technical Report, TR-ITS-2005.013 (2005)
- [11] DeLooze, L.: Attack Characterization and Intrusion Detection using an Ensemble of Self-Organizing Maps. In: *IEEE Workshop on Information Assurance United States Military Academy*, West Point, NY, pp. 108–115 (2006)
- [12] Lakhina, A., Crovella, M., Diot, C.H.: Characterization of network-wide anomalies in traffic flows. In: *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pp. 201–206 (2004)
- [13] Defense Advanced Research Projects Agency DARPA Intrusion Detection Evaluation Data Set, <http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html>
- [14] WIDE Project. MAWI Working Group Traffic Archive at tracer.csl.sony.co.jp/mawi/
- [15] The CAIDA Dataset on the Witty Worm, Colleen Shanon and David Moore (2004), <http://www.caida.org/passive/witty>
- [16] Wei, L., Ghorbani, A.: Network Anomaly Detection Based on Wavelet Analysis. *EURASIP Journal on Advances in Signal Processing* Article ID 837601, 16 (2009), doi:10.1155/2009
- [17] Troop, J.A.: Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Transactions on Information Theory* 50(10) (2004)

Stochastic Modelling of Sentence Semantics in Speech Recognition

Włodzimierz Kasprzak and Paweł Przybysz

Abstract. A stochastic approach to spoken sentence recognition is proposed for the purpose of an automatic voice-based dialogue system. Three main tasks are distinguished: word recognition, word chain filtering and sentence recognition. The first task is solved by typical acoustic processing followed by phonetic word recognition with the use of Hidden Markov Models (HMM) and Viterbi search. For the second solution an N-gram model of natural language is applied and a token-passing search is designed for the filtering of important word chains. The third task is solved due to a semantic HMM of sentences. The final sentence is recognized and a meaning is assigned to its elements with respect to given application domain. A particular spoken sentence recognition system has been implemented for train connection queries.

1 Introduction

Spoken sentence analysis is a multi-disciplinary problem in which techniques are involved that originate from signal processing, phonetics, computational linguistics and natural language processing [1]. In engineering disciplines speech processing can be decomposed into stages of acoustic-, phonetic-, lexical-, syntactic-, semantic- and pragmatic analysis. The main application of spoken sentence analysis, that we consider in this paper, are automatic spoken language dialog systems (e.g. automatic railway information system) [2]. This limits our interest to first 4 stages of speech processing and to a limited-scope semantic analysis.

General semantic analysis of natural languages usually requires large data bases. For example, research in psycholinguistics identifies how humans process a natural language. The goal of *WordNet* project [3] was to create a data base for *lexical and semantic memory*, i.e. allowing to search the dictionary by associations

Włodzimierz Kasprzak · Paweł Przybysz
Institute of Control and Computation Engineering
Warsaw University of Technology
ul.Nowowiejska 15/19, 00-665 Warszawa, Poland
e-mail: [W.Kasprzak, P.Przybysz2}@elka.pw.edu.pl](mailto:{W.Kasprzak, P.Przybysz2}@elka.pw.edu.pl)

originating from the grammar form of a word and/or semantic relations between words in given language. Its use is much broader than automatic dialogue systems - a multi-language WordNet will support automatic language translation and speech understanding (the pragmatic analysis of speech).

The studies on language *ontology* [11], among others, lead to inheritance relations between words based on their meanings. Such relations can then support the pragmatic analysis of sentences, allowing to exchange words with similar meaning depending on the context of their current use.

The syntactic-semantic analysis of spoken sentences is most often based on: case-frame grammars [6], probabilistic grammars [7] and stochastic models [10]. They seem to be useful for specific dialogue systems in which only a limited subset of a natural language need to be considered. The advantage of a case-frame grammar in comparison to context-free formal grammars is the ability to express large number of sentence configurations without a need to generate them all. The idea is to build the sentence "around" a key word which represents the main use case of a sentence.

The statistical approach to speech recognition is widely recognized. The acquired signal is contaminated by noise, the signal shape is of high variability and depends on the speaker and even the meaning of properly syntactically recognized sentences is often ambiguous. All this motivates the use of stochastic models in speech recognition.

The paper is organized as follows. Section 2 presents and discusses the structure of our system. The next section 3 concentrates on the acoustic-phonetic modelling of spoken words. In section 4 an N-gram estimation approach is presented. In section 5 the *token-passing* search is presented, with the goal to filter possible word sequences by using N-gram models. In section 6 the idea of a HMM for sentence recognition and meaning assignment is introduced. An example is presented in section 7 - the recognition of train departure questions.

2 System Structure

The speech recognition system is structured into man abstraction levels: acoustic analysis, phonetic analysis (word recognition), word chain filtering and sentence recognition. The symbolic part of sentence analysis is split into the syntax-driven detection of word sequences and a semantic-driven sentence recognition. From syntactic point of view a sentence is a chain of words, where each word is again a sequence of phonemes. From system point of view we can distinguish the need of *model creation* and of *model use* (for the purposes of word and sentence recognition).

In Fig. 1 the proposed structure of our speech analysis system is outlined. The input signal is converted to a sequence of numeric feature vectors, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$, that represent acoustic features of consecutive signal frames. This is a feature detection step.

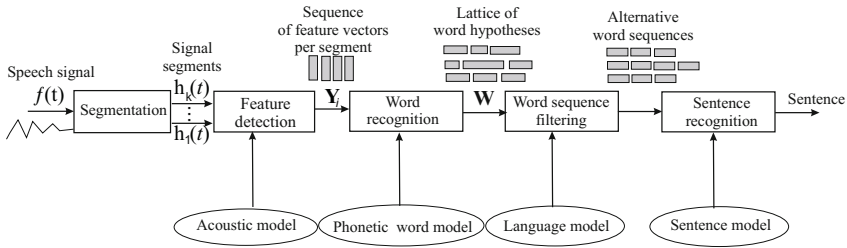


Fig. 1 The speech recognition system

The next step is to recognize a sequence of words, $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_K)$, that in a best way (with highest probability among competitive candidates induced by the language model) matches the measured sequence \mathbf{Y} .

The word sequence recognition step (in some works also called as *decoding*) solves the following stochastic task:

$$\hat{W} = \max_W [p(W|Y)] \tag{1}$$

But the distribution $p(W|Y)$ represents a diagnostic relation (from observed effect to hidden effect). Hence, it is difficult to obtain its direct model. Applying the Bayes rule we obtain:

$$p(W|Y) = \frac{p(Y|W) p(W)}{p(Y)} \tag{2}$$

that means, we express the problem in terms of prior probabilities:

$$\hat{W} = \max_W [p(Y|W)] p(Y) \tag{3}$$

For word sequence recognition we shall create two stochastic models:

1. a phonetic model - which gives the conditional probability of signal measurement for given sentence, $p(Y|W)$, and
2. a language model - which gives probabilities of word sequences, $p(W)$.

The Hidden Markow Model (HMM) [10], [12] allows for the transition both from acoustic to phonetic description and from phonemes to words. A sequence of phonetic entities representing a spoken is encoded as a sequence of HMM-states, $(s_1 \dots s_m)$, whereas the measured sequence of frame features, $(y_1 \dots y_n)$ is a sequence of observation variable values in HMM. The *Viterbi search* [8] is applied to find the best match between such two sequences, or in other words - the best path in the HMM of all words.

The required stochastic language model, $P(W) \sim P(w_1, w_2, \dots, w_n)$, should allow to select proper sentences in given language and help to reject wrongly generated

word sequences. For example, $P(\text{"W czym mogę Panu pomóc"})$ should be of high probability value if compared to $P(\text{"Jak biegać kot chodzić pies"})$.

In practice it will be difficult to learn the estimates of all individual sentences. We shall rely on shorter word sequences and learn probabilistic models, called N-grams, in which a word's probability is conditioned upon at most $N - 1$ direct predecessor words [4], [7], [12]. In this work the language model will be used by a *token-passing* search to perform a filtering of word chains.

Another HMM will be applied for sentence recognition, but it will rather represent the "semantics" of a sentence in given application. The words will be observed and accepted by HMM states due to their meaning and not their syntactic role. Thus, the HMM states represent key structures of a sentence to that allow to recognize the type of sentence (i.e. question, information, order) and to associate an interpretation (action, answer) to it.

3 Acoustic-Phonetic Model of Spoken Words

3.1 Acoustic Model

Here we apply the well-known scheme of *mel-cepstral* coefficients (MFCC) [1]. In every considered signal segment of around 16 ms duration time a vector of 38 numeric features is detected.

3.2 Phonetic Model

All words in a dictionary are given phonetic transcriptions in terms of 39 phonemes. Each phoneme is divided into 1-, 2- or 3 parts, called three-phones.

The acoustic and phonetic models of spoken words (from given dictionary) are combined into a single HMM. The phonetic model of every word is expressed by the structure of a left-to-right HMM, i.e. its states, s_i , and transition probabilities, a_{ij} , between pairs of states s_i and s_j , where s_j follows s_i .

The acoustic feature vector y may be attached to a HMM via possible three-phones and their output probabilities, λ_{jm} (for observed phone m in state j) [10], [12]. In a so called semi-continues probability model, the output probability takes the form of a Gaussian mixture, i.e. the probability of observing in state j a vector y is computed as:

$$b_j(y) = \sum_{m=1}^M \lambda_{jm} N(y; \mu_{jm}, \sigma_{jm}) \quad (4)$$

3.3 Word Hypothesis Lattice

A slightly modified *Viterbi search* [1] is applied to find a best match between every path in the HMM word model and a sequence of signal segments, that we assume to contain spoken utterances of 1-, 2- or 3-sylab words.

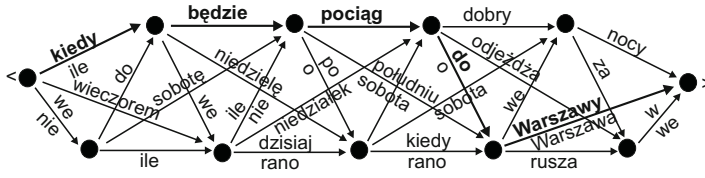


Fig. 2 Illustration of the word hypothesis lattice. Besides the optimal sequence ("Kiedy będzie pociąg do Warszawy") many other competing sequences are still possible.

The competing word sequences can be represented by a lattice-graph, where the nodes are ordered according to the (assumed) syllab (and time) index (Fig. 2). Every arc in this graph represents the best word recognition results achieved for the appropriate signal frame. More than one competing word hypothesis can be stored for a given signal frame, depending on their quality scores.

4 Language Model - N-Grams

In a formal grammar a sentence W belongs or not to some language L . Hence the probability $P(W)$ that the sentence is proper in language L is 0 or 1.

In a probabilistic context-free grammar (PCFG) [7], [12] every production rule has assigned its probability value. Hence, the probability of a sentence is a function of rule probabilities needed to generate this sequence. The key difficulty is to design this function that combines probabilities.

In opposite to PCFG-s stochastic word sequence models, called N-grams, are easy to generate and to modify. But they have to cope with the "sparse" nature of training data.

4.1 The N-Gram Model

Let $W, W = (w_1 w_2 \dots w_n)$, be a word chain. When the word chain satisfies a Markov random process of order $(N - 1)$, where $N \leq n$, its N-gram model is:

$$P(W) = \prod_{i=1}^n P(w_i | w_{i-N+1}, \dots, w_{i-1}) \tag{5}$$

In practice the N-grams are limited to $N = 2, 3, 4$ and such special cases are called [4]: unigram, $P(w_i)$, bigram, $P(w_i | w_{i-1})$, three-gram, $P(w_i | w_{i-1}, w_{i-2})$, four-gram, $P(w_i | w_{i-1}, w_{i-2}, w_{i-3})$. These distributions are estimated during a learning process that is based on the *maximum likelihood* rule (ML). The elementary learning approach is to count the appearance frequency of given word chain in the training data.

Let $C(w_{i-N+1} \dots w_{i-N+N})$ is the number of training set appearances of given word chain of length N . The unigram estimation is simple (w_k means any word in the training set):

$$P(w_i) = \frac{C(w_i)}{\sum_{w_k} w_k} \quad (6)$$

The bigram estimation needs to count the number of appearances of given pair of words and of the first word:

$$P(w_i|w_{i-1}) = \frac{C(w_{i-1}w_i)}{C(w_{i-1})} \quad (7)$$

This can be generalized for an N-gram as:

$$P(w_i|w_{i-N+1}, w_{i-N+2}, \dots, w_{i-1}) = \frac{C(w_{i-N+1}w_{i-N+2}\dots w_i)}{C(w_{i-N+1}w_{i-N+2}\dots w_{i-1})} \quad (8)$$

4.2 N-Gram Learning with Smoothing

The drawback of elementary N-gram learning approach is the sparsity of available training data, even when the number of words is large. For example in a text that contains several million of words around 50% of three-grams may appear only one time, and around 80% of three-grams - no more than 5 times. Obviously such sparse data set will lead to large N-gram estimation errors.

To eliminate such drawback a common procedure in N-gram learning is to apply a *smoothing* operation on the estimated probability distribution. The simplest smoothing method, called *Laplace smoothing*, is to add 1 to the number a sequence appeared in the training set [4].

4.3 The Katz Smoothing Method

In this work we apply the so called *Katz* method [9]:

1. for frequent N-tuples of words apply the elementary estimation approach,
2. for rare N-tuples of words apply the 'good Turing estimate',
3. for non-observed N-tuples of words apply a smoothing method by *returning to (N-1)-grams*.

The *good Turing estimate* takes the form (for a bigram):

$$C^*(w_i|w_{i-1}) = \begin{cases} r, & r > k \\ d_r r, & 0 < r \leq k \\ \alpha(w_{i-1})P(w_i), & r = 0 \end{cases} \quad (9)$$

Here r denotes the appearance number, i.e. $r = C(w_{i-1}, w_i)$, d_r - the discount rate, and α - a normalization coefficient. The parameter k , that selects one of the 3 approaches, is set by default - for example it may be set to 5.

$\alpha(w_{i-1})$ is estimated in such a way to satisfy the following condition:

$$\sum_{w_i} C^*(w_{i-1}, w_i) = \sum_{w_i} C(w_{i-1}, w_i) \quad (10)$$

The discount rate is computed as:

$$d_r = \frac{\frac{r^*}{r} - \frac{(k+1)n_{k+1}}{n_1}}{1 - \frac{(k+1)n_{k+1}}{n_1}} \quad (11)$$

for $r = 1, 2, \dots, k$. Where n_r denotes the number of N-tuples, that appear exactly r times in the training set, and r^* is the good Turing estimate, i.e.:

$$r_r^* = (r + 1) \frac{n_{r+1}}{n_r} \quad (12)$$

Thus the probabilities in a three-gram model are estimated according to the following options:

$$P^*(w_i | w_{i-1}, w_{i-2}) = \begin{cases} \frac{C(w_{i-2}w_{i-1}w_i)}{C(w_{i-2}w_{i-1})}, & r > k \\ d_r \frac{C(w_{i-2}w_{i-1}w_i)}{C(w_{i-2}w_{i-1})}, & 0 < r \leq k \\ \alpha(w_{i-1}, w_{i-2})P(w_i | w_{i-1}), & r = 0 \end{cases} \quad (13)$$

5 Token-Passing Search

We developed a general-purpose search, called *token-passing* search, in order to convert the lattice of word hypotheses into meaningful sentences of words. This is a breadth-first search controlled by evaluations of partial word sequences (paths) and which uses the N-gram model, corresponding to the dictionary of current application domain, to prune paths with inconsistent n -tuples of words. Thus the *token passing* search takes as input: 1) the *dictionary* model, represented by the integrated HMM for words, and 2) the *language* model given by N-grams.

A *token* means a data structure associated with the search tree node that contains: 1) the score (evaluation) of corresponding path, P , and 2) a pointer, *link*, to path description (a structure R).

A search tree node passes its token to its successor nodes. A new structure of type R is created for every successor node that holds: a link to obtained predecessor token, the new added word with its lattice time index and its quality score. Then the score of every new token is modified by the product of N -gram probability (of added word upon the condition of $N - 1$ predecessor words) and its quality score.

6 Semantic HMM

A sentence is a sequence of words. We propose the use of another HMM model for representation of a stochastic syntax of sentences in given application. States in this model correspond to word categories, whereas observations can be specific words given in the dictionary.

The categories of words are distinguished from point of view of the application domain. Hence, the states of such HMM represent *semantic* entities, rather than syntactic ones.

Semantic HMM is based on assumption that every sentence is combined from parts containing *atomic* semantic information. For example, an atomic part may be: a question form (*when, where, at what time*), a time period (*at eight a.m., afternoon, at evening*) or destination (*Warszawa*).

Prior probability of a valid sequence of semantic parts is expressed by transition probabilities along the appropriate path in HMM. Posterior probability is found by including the word acceptance (observation) probabilities in states along such path. Specific features of the semantic HMM are:

- It can represent a sentence category (many sentences with the same meaning) rather than a single sentence ;
- During the recognition process a particular sentence from given category is detected along with sentence meaning. Hence, speech recognition system could execute an action according to recognized sentence;
- Semantic parts can be used as a elements of many semantic HMMs.

In a general use of this approach the HMM states would be rather of syntactic meaning, i.e. they may represent:

- the role in a sentence: subject, predicate, object;
- a syntax category: noun, verb, conjunction, adjective, number, etc.

7 Example

The HMM in Fig. 3 allows a very flexible modelling of sentences. It represents the application domain: *question on train departure*. The word dictionary contains at least 38 words in base form, like: from, to, train, hour, minute, day, when, Monday, today, etc. This can be further extended by the names of cities for which train connections are needed.

Some words appear multiple times, but in different grammar forms of common base word. They are included in the word recognition stage. For simplicity of the semantic model, after the word sequence detection they are converted into the base form.

The states of HMM represent following 11 meaning (and not directly syntactic) categories: *question attribute, departure form, day, day-time, train attribute, train, from, to, departure city, destination city, end of sentence*. Every state can emit several words with specific probabilities. For example the state called *train attribute* can accept the following words: "nearest", "last", "fastest", "first".

Here are examples of valid sentences accepted by this model: *When the train from Warszawa to Krakow departs?* ("Kiedy będzie pociąg do Warszawy z Krakowa?"), *When the first train from Krakow to Warszawa departs?* ("Kiedy jest pierwszy pociąg z Krakowa do Warszawy?"), *When the train from Krakow to Warszawa departs tomorrow?* ("Kiedy jest jutro pociąg z Krakowa do Warszawy?"), *When the*

next train to Wrocław departs? ("Kiedy jest najbliższy pociąg do Wrocławia?"), When the first train from Krakow to Warszawa departs? ("O której odjeżdża pierwszy pociąg z Krakowa do Warszawy?"), When the first train from Krakow to Warszawa departs tomorrow afternoon? ("O której będzie jutro pierwszy pociąg po południu z Krakowa do Warszawy?"), When tomorrow afternoon the first train from Krakow to Warszawa departs?, ("O której jutro będzie pierwszy pociąg po południu z Krakowa do Warszawy?"), When afternoon the first train from Krakow to Warszawa departs a day after tomorrow? ("O której pojutrze będzie pierwszy pociąg po południu z Krakowa do Warszawy?"), When the train to Warszawa departs? ("O której jest pociąg do Warszawy?"), When tomorrow the train to Grodzisk leaves? ("O której jutro odjeżdża pociąg do Grodziska?").

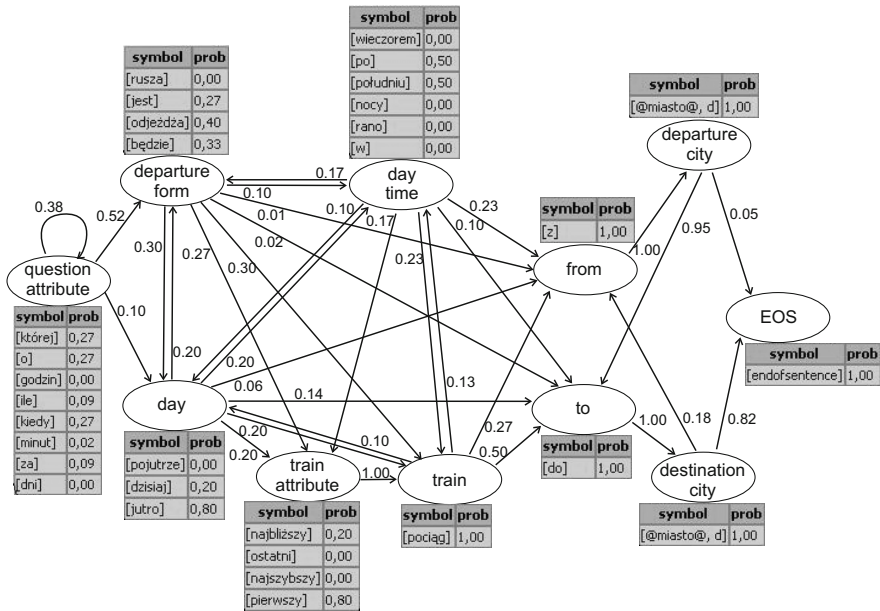


Fig. 3 Illustration of a semantic HMM for train departure information.

8 Summary

A design of a spoken sentence recognition system applied for the purpose of man-machine dialogue systems was proposed. The novel elements can be characterized as follows:

1. the stochastic modelling of words and sentences in given language by means of HMMs and N-gram models;
2. a token-passing search as a word filtering stage in this system;
3. a semantic HMM for spoken sentence recognition and meaning (action) association in some application domain (e.g. train departure information).

Acknowledgements. This work was supported by the Polish Ministry of Science and Higher Education by the grant N N514 1287 33.

References

- [1] Benesty, J., Sondhi, M.M., Huang, Y. (eds.): Springer Handbook of Speech Processing. Springer, Berlin (2008)
- [2] Bennacef, S., Devillers, L., Rosset, S., Lamel, L.: Proceedings of International Conference on Spoken Language Processing, ICSLP 1996, pp. 550–553 (1996)
- [3] Chen, S.F., Goodman, J.: Computer Speech and Language 13, 359–393 (1999)
- [4] Chen, S.F., Rosenfeld, R.: IEEE Trans. on Speech and Audio Processing 8(1), 37–50 (2000)
- [5] Fellbaum, C. (ed.): WordNet. An Electronic Lexical Database. The MIT Press, Cambridge (1998)
- [6] Hayes, P.J., Andersen, P.M., Safier, S.: Proceedings of 23rd Annual Meeting of ACL, Chicago, Illinois, pp. 153–160 (1985)
- [7] Jelinek, F., Lafferty, J.D., Mercer, R.L.: Basic methods of probabilistic context-free grammars. In: Laface, P., De Mori, R. (eds.) Speech Recognition and Understanding: Recent Advances, Trends, and Applications, pp. 345–360. Springer, Berlin (1992)
- [8] Kasprzak, W.: Rozpoznawanie obrazów i sygnałów mowy. Warsaw University of Technology Press, Warszawa (2009)
- [9] Katz, S.M.: IEEE Trans. Acoustics, Speech and Signal Proc. ASSP 35, 400–401 (1987)
- [10] Rabiner, L., Juang, B.H.: Fundamentals of Speech Recognition. Prentice Hall, New York (1993)
- [11] Russell, S., Norvig, P.: Artificial Intelligence. A Modern Approach. Prentice Hall, New York (2002)
- [12] Young, S.: HMMs and related speech recognition technologies. In: Benesty, J., Sondhi, M.M., Huang, Y. (eds.) Springer Handbook of Speech Processing, pp. 539–555. Springer, Berlin (2008)

Thai Character Recognition Using "Snakecut" Technique

Thakerng Wongsirichot, Piyawan Seekaew, and Parunyu Arnon

Abstract. Optical character recognition has become one of the powerful tools in the area of pattern recognition and artificial intelligence. Our research targets practically to Thai character recognition on Thai national identification cards. We propose a Thai character recognition technique, so called "Snakecut". It is invented by iterative investigations of the Thai characters. The "Snakecut", an unsupervised technique, is able to trace around each character, meticulously. A set of 26 weighted parameters are employed to further tuning recognition performances. Experiment results accuracy rates are 83.10% for pure Thai characters and 92.96% for a mixture of numbers and Thai characters with some limitations.

1 Introduction

Optical character recognition (OCR) is one of the most successful applications in the field of pattern recognition. OCR becomes an essential tool for many businesses such as publishing companies. [1] The OCR research area on text can be categorized into two main groups namely printed and handwritten text recognitions. [6] A number of researchers present various techniques including an extraction method of edge information from gray scale documents, a method of edge detection in order to localizing text area. [1] Others include some well-known models such as Hidden Markov Models (HMM) for the purpose of character recognitions. [5] Many practical OCRs have been constructed based on particular languages such as Telugu script. [3] Dilemmas, which are possibly encountered in OCR processing, divides into two types. Firstly, an internal OCR problem includes performance of OCRs algorithms and techniques. Secondly, external OCR problems or environmental problems include insufficient light, quality of original sources, etc. In terms of

Thakerng Wongsirichot · Piyawan Seekaew · Parunyu Arnon
Information and Communication Technology Programme,
Faculty of Science, Prince of Songkla University, Songkla, Thailand
e-mail: thakerng.w@psu.ac.th

performance comparisons are limited due to each experimental testing environments. Specially, most researches were conducted based on particular circumstances. However, previous researches are able to be used as encompasses for further researches, which are potentially advance. Optical character recognition techniques have been employed in many languages for over a decade. With the nature of each languages characters, the OCRs are required accordingly to adapt in order to achieve the maximum accuracy and robustness. In the case of Thai characters, there are a number of patterns that can be used as guidelines for detecting character correctly.

2 Principles of Thai Characters

Originally, Thai characters have been invented in 1283 by the King Ramkhamhaeng the Great. The alphabets are derived from elder scripts such as old Khmer, Tamil, Sanskrit, Malayalam scripts, etc. With the combination of scripts, Thai characters have been gradually adapted since its origin. The current Thai scripts contain mainly 44 alphabets, 32 vowels, 4 tone marks and 10 Thai numbers. A set of some Thai alphabets is illustrated in Fig. 1.

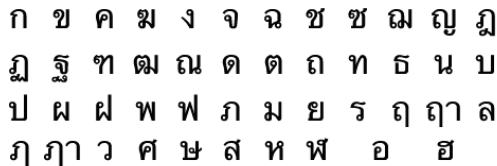


Fig. 1 A set of Thai alphabets

Thai scripts are written in standardized 4 levels namely, a main level (level 2), a lower line (level 1), and an upper level (level 3) and finally an uppermost level (level 4) as shown in Fig. 2. Thai words and sentences are commonly written or typed without any symbols such as periods for stopping words and sentences. Occasionally, some symbols such as commas and semicolons are added after some words for the purpose of distinctive clarification of words.

According to standard typed Thai characters, each Thai alphabet has its distinctive appearances. For example, the first two respective Thai alphabets are Kor Kai and Khor Khai as shown in Fig. 3.

The obvious distinction between the first two alphabets above is the headed circle. The headed circle appears in the second alphabet however the first alphabet has no headed circle. This phenomenon is one of the distinctions that assists in differentiate two of the Thai alphabets, which are Kor Kai and Tor Tong from the rest of the alphabets. Some of Thai alphabets are closely resemble therefore they cause the difficulty of character recognitions.



Fig. 2 Levels of Thai scripts



Fig. 3 Level of Thai scripts

3 Algorithms and Techniques

Due to various styles of typed Thai characters, Thai national identification cards have been touched into our attention as a standard case. Thai national identification cards have been formally introduced since 1983. The bill of national identification cards (1983) states all Thai citizens who are 15 years old but not over 70 years old must acquire national identification cards. The latest version (5th version) as shown in Fig. 4 of the cards is selected as test cases for our research [8].



Fig. 4 A sample of Thai National Identification card

There are four zones in the card are studied in this research. $zone_a$ is the area of a card owners photo. $zone_b$ is the area of an identification number located. $zone_c$ is the area in which a first name and a last name are presented. Finally, $zone_d$ is the area in which a card owner’s date of birth is presented.

Main essential algorithms of our research contain (1) Image Transformation (2) Number Separation (3) Sneakcut (4) Alphanumerical Character Recognition.

1) Image Transformation: Images of Thai national identification cards are captured via an ordinary web cam. The bitmap images are in the size of 640 by 480 pixels.

Algorithm 1. Image Transformation

Require: BMP is a set of pixels of a bitmap image of Thai national identification cards

```

1: for all  $Px_i \in BMP$  do
2:   if  $0 \leq Px_i \leq 80$  then
3:      $Px_i \leftarrow 0$ 
4:   else
5:      $Px_i \leftarrow 255$ 
6:   end if
7: end for
8: return coordinates of  $zone_a, zone_b, zone_c,$  and  $zone_d$ 

```

2) Number Separation: This proposed technique is applied to the $zone_c$, which contains only 13 digits of an identification number.

Algorithm 2. Number Separation

Require: $zone_b$ is a set of pixels in the identification number.**Require:** $BLACK$ is a set of black pixels**Require:** $IDENT(x)$ is clustering function

```

1:  $BLACK = \phi$ 
2: for all  $Px_i \in zone_b$  do
3:   if  $Px_i = 0$  then
4:      $BLACK = BLACK \cup \{Px_i\}$ 
5:   end if
6: end for
7: for all  $bpx \in BLACK$  order by ascending do
8:    $N \leftarrow IDENT(bpx)$ 
9:    $zone_{bN} = zone_{bN} \cup \{bpx\}$ 
10: end for
11: return  $zone_{bN}, 0 \leq N \leq 12$ 

```

3) Snakecut: This alphanumerical character separation technique so called "Snakecut" is applied to the $zone_c$ and $zone_d$, which contains both numbers and Thai alphabets. The Snakecut technique is imitated snake movements, which can move in any directions even in a very narrow space. It acquires an entire line of alphanumerical characters and neatly separates them into a set of single character images.

4) Alphanumerical Character Recognition: Due to the complexity of Thai alphabets, researchers proposed various techniques such as Thai Type Style Recognition, which employed Neural Network technique into recognition process[7]. However, each technique contains some trivial to moderate imperfect recognition performance. We propose a new technique which is created specially for Thai alphabets. 26 weighted parameters have been created as shown in Fig 5

Apparently, the weighted parameters from 1 to 5 indicate coordinates of the first black pixel detection from top to bottom. The weighted parameters from 6 to 10 indicate coordinates in vertically reverse order from the weighted parameters from 1 to 5. Correspondingly, the weighted parameters from 11 to 15 denote coordinates of the first black pixel detection from left to right. The weighted parameters from 16 to 20 are in the horizontally reverse order to the weighted parameters from 11 to 15. Finally, the weighted parameters from 21 to 26 are as follows.

Our research has conducted a set of preliminary studies. The weighted parameters are differently influences the character recognition performance. In order to find the

Algorithm 3. Snakecut

Require: *IMG* is an image of the entire line of alphanumerical characters

Require: *BLACK* is a set of black pixel coordinates

Require: *CUTOFF* is a set of white pixel coordinates that is located by only 1 pixel from a black pixel

Require: *IDENT*(*x*) is a clustering function

```

1:  $x \leftarrow 0$ 
2:  $y \leftarrow 0$ 
3: for all  $c(x, y) \in IMG$  do
4:   if  $c(x, y) \neq 0$  then
5:     if  $y > Max_y$  then
6:        $x++$ 
7:        $y \leftarrow 0$ 
8:     end if
9:   else
10:     $BLACK = BLACK \cup \{c(x, y)\}$ 
11:  end if
12:   $y++$ 
13: end for
14: for all  $c(x, y) \in BLACK$  order by ascending do
15:    $y' \leftarrow y$ 
16:    $y++$ 
17:    $CUTOFF \leftarrow CUTOFF \cup \{c(x, y-1)\}$ 
18:   if  $(y' < y) \cap (y \leq Max_y) \cap (y+5 \neq 0)$  then
19:      $CUTOFF = CUTOFF \cup \{c(x, y-1)\}$ 
20:   end if
21:   if  $y > Max_y$  then
22:      $x++$ 
23:      $y \leftarrow 0$ 
24:   end if
25: end for
26: for all  $c(x, y) \in CUTOFF$  order by ascending do
27:    $N \leftarrow IDEN(c(x, y))$ 
28:    $IMG_N = IMG_N \cup \{c(x, y)\}$ 
29: end for
30: return  $IMG_N$ 

```

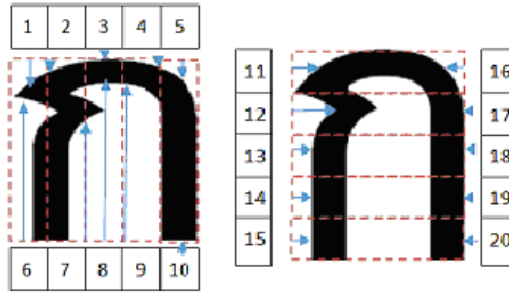


Fig. 5 26 Weighted parameters for Thai character recognition

Table 1 The weighted parameters from 21 to 26

Weighted parameter	Description
21	A number of vertical approximated black line in a character
22	A flag value of long tailed character
23	A flag value of an appearance of vowel in a lower line
24	A zone indicator ($zone_a$, $zone_b$, $zone_c$ or $zone_d$)
25	A ratio of a characters height to a main level in a line
26	A ratio of width and height of a character

best weighted parameters with a minimum error, a set of iterative testing process must be performed.

$$Min_{error} = \sum_{N=1}^{26} weight_N \times (base_N - input_N)^2 \tag{1}$$

where

$weight_N$ is a weighted parameter for each character

$base_N$ is each based character

$input_N$ is each input character

Table 2 represents the final set of weighted parameters as per a set of iterative testing processes. The most affected weighted parameters are the weighted number 24, 22, 23, and 25, respectively.

4 Experimental Tests and Results

An experimental test has been designed and conducted by processing 20 unseen national identification cards. The performance of this proposed technique is measured by accuracy rate.[2] The overall accuracy of recognition regardless of zones

Table 2 The final set of weighted parameters

Weighted Number	Weighted Parameter	Weighted Number	Weighted Parameter
1	1	14	2
2	2	15	1
3	7	16	1
4	3	17	2
5	1	18	3
6	1	19	2
7	2	20	1
8	3	21	3
9	2	22	20
10	1	23	20
11	1	24	30
12	2	25	10
13	3	26	2

is 90.91%. Specifically, the experimental result illustrates an excellent performance in $zone_b$, which is the 13 national identification numbers as shown in Table 3. In $zone_c$, the full name area, the accuracy of the recognition is 83.10%. Finally, the $zone_d$, the date of birth area, the accuracy of the recognition is 92.96%. In our test cases, the incorrect recognized characters are caused from many reasons including missing some characters, adding more characters, mistaken characters, and insufficient light. According to the result, our technique performs slightly pleasant with numbers when comparing with Thai alphabets. However, our unsupervised technique can perform up to 83.10% accuracy in the full name area, which contains solely Thai alphabets. Some other researches reached up to 95% accuracy but with the artificial neural network supervised technique.

$$Accuracy = 100x \frac{\sum_{i=1}^N 1 - x_i}{N} \tag{2}$$

where

- x_i is the i^{th} frequency of incorreced recognised characters or numbers
- N is a total number of test case.

5 Limitation and Future Work

Thai character recognitions are still evolving due to its complexity. Each technique has its own pros and cons. Our technique is outstanding in the unsupervised technique group. However, there are a number of limitations to our work.

Table 3 An experimental result

test case	A number of detected (number/characters)			A number of incorrect recognized (number/characters)		
	<i>zone_b</i>	<i>zone_c</i>	<i>zone_d</i>	<i>zone_b</i>	<i>zone_c</i>	<i>zone_d</i>
1	13	20	10	0	3	1
2	13	19	10	0	7	0
3	13	18	9	0	2	0
4	13	15	10	0	4	2
5	13	15	10	0	3	0
6	13	15	10	0	3	2
7	13	21	10	0	5	1
8	13	19	10	0	2	1
9	13	19	10	0	1	0
10	13	17	10	0	2	0
11	13	19	10	0	1	0
12	13	17	10	0	1	0
13	13	18	11	0	3	0
14	13	17	10	0	2	2
15	13	17	10	0	2	0
16	13	20	9	0	4	0
17	13	16	9	0	2	3
18	13	16	11	0	4	0
19	13	16	10	0	1	2
20	13	21	10	0	8	0

1. Font type: Our technique is only applied to Thai national identification cards. The cards contain only one font type. In our future work, various font types will be added into our test cases. Possibly, the sneakcut algorithm will require some adjustments.
2. Watermarks: Recently, all Thai national identification cards are watermarked, which can cause a problem in capturing images.
3. Insufficient light: Insufficient light is a common problem in pattern recognition. In our research, we have built a prototype of reading box. A small fluorescent light has been installed inside the box. Therefore, the image of cards can be clearly captured. In our test cases, there is only one card (No. 12) that has a problem of insufficient light.
4. Fixed size card: Our experiments are conducted based on the Thai national identification cards only. Currently, other cards are not possible due to the position of texts on cards.

6 Conclusion

Optical character recognitions are continuously developed and improved including Thai character recognitions. Our research is inventing an unsupervised technique so called "Snakecut". With the snakecut technique, a set of 26 weighted parameters assist to compromise the complication of Thai scripts. The overall accuracy of the technique is 90.91%. Particularly, the accuracy are 83.10% for pure Thai characters and 92.96% a mixture of numbers and Thai characters. Some limitations are encountered such as watermarked cards and insufficient light.

References

- [1] Alata, M., Al-shabi, M.: Text detection and character recognition using fuzzy image processing. *Journal of Electrical Engineering* 57, 258–267 (2006)
- [2] Baeza-Yates, R., Ribeiro-Neto, B.: *Modern Information Retrieval*. Addison Wesley, Reading (1999)
- [3] Hara, S.: Ocr for japanese classical documents segmentation of cursive characters. In: PNC 2004 Annual Conference on Conjunction with PRDLA, National Institute of Japanese (2004)
- [4] Kawtrakul, A., Waewsawangwong, P.: Multi-feature extraction for printed thai character recognition. In: *Fourth Symposium on Natural Language Processing*, pp. 1–8 (2000)
- [5] Krstovski, K., MacRostie, E., Prasad, R., Natarajan, P.: End-to-end trainable thai ocr system using hidden markov models. In: *IAPR International Workshop on Document Analysis Systems*, vol. 0, pp. 607–614 (2008)
- [6] Mitranont, J., Kiwprasopsak, S.: The development of the feature extraction algorithms for thai handwritten character recognition system. In: Henttlass, T., Ali, M. (eds.) *IEA/AIE 2002. LNCS (LNAI)*, vol. 2358, pp. 77–89. Springer, Heidelberg (2002)
- [7] Tanprasert, C., Sae-Tang, S.: Thai type style recognition. In: *IEEE International Symposium on Circuit and Systems*, pp. 336–339 (1999)
- [8] The Ministry of Interior Department of Provincial Administration, Thai national identification cards
- [9] Wyatt, D.K.: *A Short History of Thailand*. Yale University Press, New Haven (2003)

Regularized NNLS Algorithms for Nonnegative Matrix Factorization with Application to Text Document Clustering

Rafal Zdunek

Abstract. Nonnegative Matrix Factorization (NMF) has recently received much attention both in an algorithmic aspect as well as in applications. Text document clustering and supervised classification are important applications of NMF. Various types of numerical optimization algorithms have been proposed for NMF, which includes multiplicative, projected gradient descent, alternating least squares and active-set ones. In this paper, we discuss the selected Non-Negatively constrained Least Squares (NNLS) algorithms (a family of the NNLS algorithm proposed by Lawson and Hanson) that belong to a class of active-set methods. We noticed that applying the NNLS algorithm to the Tikhonov regularized LS objective function with a regularization parameter exponentially decreasing considerably increases the accuracy of data clustering as well as it reduces the risk of getting stuck into unfavorable local minima. Moreover, the experiments demonstrate that the regularized NNLS algorithm is superior to many well-known NMF algorithms used for text document clustering.

1 Introduction

NMF decomposes an input matrix into lower-rank factors that have nonnegative values and usually some physical meaning or interpretation. The reduced dimensionality is very useful in data analysis due to many advantages such as denoising, computational efficiency, greater interpretability and easier visualization. Hence, NMF has already found diverse applications [2, 4, 7, 11, 12, 18, 20, 21, 22, 23, 24] in multivariate data analysis, machine learning, and signal/image processing.

Text document clustering can be regarded as the unsupervised classification of documents into groups (clusters) that have similar semantic features. The

Rafal Zdunek

Institute of Telecommunications, Teleinformatics and Acoustics,

Wroclaw University of Technology,

Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

e-mail: rafal.zdunek@pwr.wroc.pl

grouping can be basically obtained with hierarchical (tree-like structure) or partitional (flat clustering) techniques [13]. Xu *et al* [24] showed that NMF can perform text document partitioning by capturing semantic features in a document collection and grouping the documents according to similarity of these features. Different from Latent Semantic Indexing (LSI), NMF provides nonnegative semantic vectors that do not have to be orthogonal or even independent. Moreover, the coefficients of linear combinations of the semantic vectors are also nonnegative, which means that only additive combinations are permitted. The semantic vectors in NMF have also easier interpretation - namely each of them represents a base topic. Thus, each document can be represented by an additive combination of the base topics. Partitional clustering can be readily performed in the space of linear combination coefficients.

An intensive development of NMF-based techniques for document clustering brought in a variety of nonnegativity constrained optimization algorithms. Xu *et al* [24] applied the standard and weighted Lee-Seung multiplicative algorithms to document clustering, and demonstrated that NMF outperforms SVD-based clustering methods. Shahnaz *et al* [22] proposed the GD-CLS algorithm that combines the standard multiplicative algorithm for updating the semantic vectors and the regularized projected LS algorithm with a constant regularization parameter for updating the other factor. Several NMF algorithms (tri-NMF, semi-orthogonal and convex NMF) for partitional clustering have been developed by Ding *et al* [8, 9, 10, 19]. Another group of NMF algorithms is based on so-called Graph NMF (GNMF) [6] that incorporates a priori information on a local structure of clustered data. Locality Preserving NMF (LPNMF), which has been proposed by Cai *et al* [5] for document clustering, is an extension of GNMF.

The above-mentioned algorithms mostly perform multiplicative updates that assure a monotonic convergence but their convergence rate is usually very small. To accelerate the slow convergence, several additive update algorithms have been proposed, including Projected Gradient (PG) descent, Alternating Least Squares (ALS) and active-set algorithms. A survey of PG and ALS algorithms for NMF can be found in [7, 25]. Since the estimated factors in NMF-based document clustering are expected to be large and very sparse, a good choice seems to be active-set algorithms.

In this paper, we discuss the selected active-set algorithms that are inspired by the Non-Negative Least Squares (NNLS) algorithm, proposed by Lawson and Hanson [17] in 1974. Bro and de Jong [3] considerably accelerate the NNLS algorithm by rearranging computations for cross-product matrices. The solution given by the NNLS algorithms is proved to be optimal according to the Karush-Kuhn-Tucker (KKT) conditions. Unfortunately, these algorithms are not very efficient for solving nonnegativity constrained linear systems with multiple Right-Hand Side (RHS) vectors since they compute a complete pseudoinverse once for each RHS. To tackle this problem, Bentham and Keenan [1] devised the Fast Combinatorial NNLS (FC-NNLS) algorithm and experimentally demonstrated it works efficiently for energy-dispersive X-ray spectroscopy data. The FC-NNLS algorithm works also efficiently for some NMF problems. Kim and Park [16] applied the FC-NNLS algorithm to

the l_1 - and l_2 -norm regularized LS problems in NMF, and showed that the regularized NNLS algorithms work very efficiently for gene expression microarrays. Their approach assumes constant regularization parameters to enforce the desired degree of sparsity.

Here we apply the FC-NNLS algorithm to text document clustering, assuming the regularization parameter decrease exponentially with iterations to trigger a given character of iterative updates.

The paper is organized in the following way. The next section discusses the concept of NMF for document clustering. Section 3 is concerned with the NNLS algorithms. The experiments for text document clustering are presented in Section 4. Finally, the conclusions are given in Section 5.

2 NMF for Document Clustering

According to the model of document representation introduced in [24], first the collection of documents to be clustered is subject to stop-words removal and words stemming, and then the whole processed collection is represented by the term-document matrix $\mathbf{Y} = [y_{it}] \in \mathbb{R}_+^{I \times T}$, where T is the number of documents in the document collection, I is the number of words after preprocessing of the documents, and \mathbb{R}_+ is the nonnegative orthant of the Euclidean space \mathbb{R} . Each entry of \mathbf{Y} is modeled by:

$$y_{it} = t_{it} \log \left(\frac{T}{d_i} \right), \tag{1}$$

where t_{it} denotes the frequency of the i -th word in the t -th document, and d_i is the number of documents containing the i -th word. Additionally, each column vector of \mathbf{Y} is normalized to the unit l_2 -norm. The matrix \mathbf{Y} is very sparse, since each document contains only a small portion of the words that occur in the whole collection of documents.

The aim of NMF is to find such lower-rank nonnegative matrices $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{X} \in \mathbb{R}^{J \times T}$ that $\mathbf{Y} \cong \mathbf{AX} \in \mathbb{R}^{I \times T}$, given the matrix \mathbf{Y} , the lower rank J (the number of topics), and possibly a prior knowledge on the matrices \mathbf{A} and \mathbf{X} . Assuming each column vector of $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_T]$ represents a single observation (a datum point in \mathbb{R}^I), and J is *a priori* known number of clusters (topics), we can interpret the column vectors of $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_J]$ as the semantic feature vectors or centroids (indicating the directions of central points of clusters in \mathbb{R}^I) and the entry x_{jt} in $\mathbf{X} = [x_{jt}]$ as an indicator (probability) how the t -th document is related to the j -th topic.

To estimate the matrices \mathbf{A} and \mathbf{X} from \mathbf{Y} , we assume the squared Euclidean distance:

$$D(\mathbf{Y}||\mathbf{AX}) = \frac{1}{2} \|\mathbf{Y} - \mathbf{AX}\|_F^2. \tag{2}$$

The gradient matrices of (2) with respect to \mathbf{A} and \mathbf{X} are expressed by:

$$\mathbf{G}_A = [g_{ij}^{(A)}] = \nabla_A D(\mathbf{Y}||\mathbf{AX}) = (\mathbf{AX} - \mathbf{Y})\mathbf{X}^T \in \mathbb{R}^{I \times J}, \tag{3}$$

$$\mathbf{G}_X = [g_{jt}^{(X)}] = \nabla_X D(\mathbf{Y}|\mathbf{A}\mathbf{X}) = \mathbf{A}^T (\mathbf{A}\mathbf{X} - \mathbf{Y}) \in \mathbb{R}^{J \times T}. \quad (4)$$

From the stationarity of the objective function, we have $\mathbf{G}_A \triangleq 0$ and $\mathbf{G}_X \triangleq 0$, which leads to the ALS algorithm:

$$\mathbf{A} \leftarrow \mathbf{Y}\mathbf{X}^T (\mathbf{X}\mathbf{X}^T)^{-1}, \quad \mathbf{X} \leftarrow (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}. \quad (5)$$

The update rules (5) do not guarantee nonnegative solutions which are essential for NMF. Some operations are necessary to enforce nonnegative entries. The simplest approach is to replace negative entries with zero-values, which leads to the projected ALS algorithm:

$$\mathbf{A} \leftarrow \mathcal{P}_{\Omega_A} [\mathbf{Y}\mathbf{X}^T (\mathbf{X}\mathbf{X}^T)^{-1}], \quad \mathbf{X} \leftarrow \mathcal{P}_{\Omega_X} [(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}], \quad (6)$$

where

$$\Omega_A = \{\mathbf{A} \in \mathbb{R}^{I \times J} : a_{ij} \geq 0\}, \quad \Omega_X = \{\mathbf{X} \in \mathbb{R}^{J \times T} : x_{jt} \geq 0\}. \quad (7)$$

Unfortunately, the projected ALS algorithm does not assure that the limit point is optimal in the sense it satisfies the KKT optimality conditions:

$$g_{ij}^{(A)} = 0, \text{ if } a_{ij} > 0, \quad \text{and} \quad g_{ij}^{(A)} > 0, \text{ if } a_{ij} = 0. \quad (8)$$

$$g_{jt}^{(X)} = 0, \text{ if } x_{jt} > 0, \quad \text{and} \quad g_{jt}^{(X)} > 0, \text{ if } x_{jt} = 0, \quad (9)$$

and the complementary slackness:

$$g_{ij}^{(A)} a_{ij} = 0, \quad g_{jt}^{(X)} x_{jt} = 0. \quad (10)$$

3 NNLS Algorithms

Several update rules have been proposed in the literature that satisfy the KKT conditions. Here we restrict our considerations to the NNLS algorithms for updating the matrix \mathbf{X} , given the matrices \mathbf{A} and \mathbf{Y} . The discussed algorithms can also update the matrix \mathbf{A} by applying them to the transposed system: $\mathbf{X}^T \mathbf{A}^T = \mathbf{Y}^T$.

3.1 NNLS Algorithm for Single RHS

The NNLS algorithm was originally proposed by Lawson and Hanson [17], and currently it is used by the command `lsqnonneg(.)` in Matlab. This algorithm (denoted by LH-NNLS) iteratively partitions the unknown variables into the passive set P that contains basic variables and the active set R that contains active constraints, and updates only the basic variables until the complementary slackness condition in (10) is met. Let $P = \{j : x_{jt} > 0\}$ and $R = \{1, \dots, J\} \setminus P$, and the partition: $\forall t : \mathbf{x}_t = [\mathbf{x}_t^{(P)}; \mathbf{x}_t^{(R)}]^T \in \mathbb{R}^J$, and $\mathbf{g}_t = \nabla_{x_t} D(\mathbf{y}_t | \mathbf{A}\mathbf{x}_t) = [\mathbf{g}_t^{(P)}; \mathbf{g}_t^{(R)}]^T \in \mathbb{R}^J$.

The columns of \mathbf{A} can be also partitioned in the similar way: $\mathbf{A} = [\mathbf{A}_P \mathbf{A}_R]$, where $\mathbf{A}_P = [a_{*,P}]$ and $\mathbf{A}_R = [a_{*,R}]$. The basic variables can be estimated by solving the unconstrained LS problem:

$$\min_{\mathbf{x}_t^{(P)}} \left\{ \|\mathbf{A}_P \mathbf{x}_t^{(P)} - \mathbf{y}_t\|_2 \right\}, \tag{11}$$

where \mathbf{A}_P has full column rank. The nonbasic variables in the KKT optimality point should be equal zero. In contrary to the projected ALS, the NNLS algorithm does not replace negative entries with zero-values, which is equivalent to determine nonbasic variables from unconstrained LS updates, but it starts from all nonbasic variables and tries to iteratively update a set of basic variables. The LH-NNLS algorithm is given by Algorithm 1.

Algorithm 1: LH-NNLS Algorithm

Input : $\mathbf{A} \in \mathbb{R}^{I \times J}$, $\mathbf{y}_t \in \mathbb{R}^I$

Output: $\mathbf{x}_t^* \geq 0$ such that $\mathbf{x}_t^* = \arg \min_{\mathbf{x}_t} \|\mathbf{A}\mathbf{x}_t - \mathbf{y}_t\|_2$

```

1 Initialization:  $P = \emptyset$ ,  $R = \{1, \dots, J\}$ ,  $\mathbf{x}_t = \mathbf{0}$ ,  $\mathbf{g}_t = -\mathbf{A}^T \mathbf{y}_t$ ;
2 repeat
3    $k \leftarrow k + 1$ ;
4    $m = \arg \min_{j \in R} \{g_{jt}\}$ ; // the constraint to add
5   if  $g_{mt} < 0$  then
6      $P \leftarrow P \cup m$ , and  $R \leftarrow R \setminus m$ ;
7   else
8     stop with  $\mathbf{x}_t$  as an optimal solution
9    $\bar{\mathbf{x}}_t^{(P)} = ((\mathbf{A}_P)^T \mathbf{A}_P)^{-1} (\mathbf{A}_P)^T \mathbf{y}_t$  where  $\mathbf{A}_P = [a_{*,P}] \in \mathbb{R}^{I \times |P|}$ ;
10  while  $\bar{x}_{jt}^{(P)} \leq 0$  for  $j = 1, \dots, J$  do
11     $\alpha = \min_{\substack{j \in P \\ \bar{x}_{jt}^{(P)} \leq 0}} \left\{ \frac{x_{jt}^{(P)}}{x_{jt}^{(P)} - \bar{x}_{jt}^{(P)}} \right\}$ ; // the step length
12     $\mathbf{x}_t \leftarrow \left[ \mathbf{x}_t^{(P)} + \alpha (\bar{\mathbf{x}}_t^{(P)} - \mathbf{x}_t^{(P)}); \mathbf{0} \right]^T$ ;
13     $N = \{j : x_{jt} = 0\}$ ; // the constraint to drop
14     $P \leftarrow P \setminus N$ , and  $R \leftarrow R \cup N$ ;
15     $\bar{\mathbf{x}}_t^{(P)} = ((\mathbf{A}_P)^T \mathbf{A}_P)^{-1} (\mathbf{A}_P)^T \mathbf{y}_t$  where  $\mathbf{A}_P = [a_{*,P}] \in \mathbb{R}^{I \times |P|}$ ;
16     $\mathbf{x}_t \leftarrow \left[ \bar{\mathbf{x}}_t^{(P)}; \mathbf{0} \right]^T$ ;
17 until  $R = \emptyset$  or  $\max_{j \in R} \{g_{jt}\} < tol$ ;

```

Bro and de Jong [3] considerably speed up this algorithm for $I \gg J$ by precomputing the normal matrix $\mathbf{A}^T \mathbf{A}$ and the vector $\mathbf{A}^T \mathbf{y}_t$, and then replace the steps 9 and 15 in Algorithm 1 with:

$$\bar{\mathbf{x}}_t^{(P)} = ((\mathbf{A}^T \mathbf{A})_{P,P})^{-1} (\mathbf{A}^T \mathbf{y}_t)_P. \quad (12)$$

Unfortunately, the inverse of $(\mathbf{A}^T \mathbf{A})_{P,P}$ must be computed for each t , which is very expensive if the number of RHS is very large.

3.2 Regularized NNLS Algorithm for Multiple RHS

Van Benthem and Keenan [11] tackled the problem of a high computational cost for multiple RHS. They noticed that for a sparse solution with multiple column vectors, a probability of finding columns vectors that have the same layout of the zero-entries (active constraints) is high. Hence, after detecting such vectors in \mathbf{X} , their passive entries can be updated computing the inverse of $(\mathbf{A}^T \mathbf{A})_{P,P}$ only once. The NNLS algorithm proposed by Van Benthem and Keenan is referred to as FC-NNLS. The CSSLS algorithm (Algorithm 3) is a part of the FC-NNLS algorithm.

We applied the FC-NNLS algorithm to the penalized Euclidean function:

Algorithm 2: Regularized FC-NNLS Algorithm

Input : $\mathbf{A} \in \mathbb{R}^{I \times J}$, $\mathbf{Y} \in \mathbb{R}^{I \times T}$, $\lambda \in \mathbb{R}_+$

Output: $\mathbf{X}^* \geq 0$ such that $\mathbf{X}^* = \arg \min_{\mathbf{X}} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_F + \lambda \|\mathbf{X}\|_F$

1 **Initialization**: $M = \{1, \dots, T\}$, $N = \{1, \dots, J\}$;

2 Precompute: $\mathbf{B} = [b_{ij}] = \mathbf{A}^T \mathbf{A} + \lambda \mathbf{I}_J$ and $\mathbf{C} = [c_{it}] = \mathbf{A}^T \mathbf{Y}$;

3 $\mathbf{X} = \mathbf{B}^{-1} \mathbf{C}$; // unconstrained minimizer

4 $\mathbf{P} = [p_{jt}]$, where $p_{jt} = \begin{cases} 1 & \text{if } x_{jt} > 0, \\ 0 & \text{otherwise} \end{cases}$; // passive entries

5 $F = \{t \in M : \sum_j p_{jt} \neq 1\}$; // set of columns to be optimized

6 $x_{jt} \leftarrow \begin{cases} x_{jt} & \text{if } p_{jt} = 1, \\ 0 & \text{otherwise} \end{cases}$;

7 **while** $F \neq \emptyset$ **do**

8 $\mathbf{P}_F = [p_{*,F}] \in \mathbb{R}^{J \times |F|}$, $\mathbf{C}_F = [c_{*,F}] \in \mathbb{R}^{J \times |F|}$;

9 $[x_{*,F}] = \text{cssls}(\mathbf{B}, \mathbf{C}_F, \mathbf{P}_F)$; // Solved with the CSSLS algorithm

10 $H = \{t \in F : \min_{j \in N} \{x_{jt}\} < 0\}$; // Columns with negative vars.

11 **while** $H \neq \emptyset$ **do**

12 $\forall s \in H$, select the variables to move out of the passive set \mathbf{P} ;

13 $\mathbf{P}_H = [p_{*,H}] \in \mathbb{R}^{J \times |H|}$, $\mathbf{C}_H = [c_{*,H}] \in \mathbb{R}^{J \times |H|}$;

14 $[x_{*,H}] = \text{cssls}(\mathbf{B}, \mathbf{C}_H, \mathbf{P}_H)$;

15 $H = \{t \in F : \min_{j \in N} \{x_{jt}\} < 0\}$; // Columns with negative vars.

16 $\mathbf{W} = [w_{jt}] = \mathbf{C}_F - \mathbf{B}\mathbf{X}_F$, where $\mathbf{X}_F = [x_{*,F}]$; // negative gradient

17 $Z = \{t \in F : \sum_j w_{jt}(1 - \mathbf{P}_F)_{jt} = 0\}$; // set of optimized columns

18 $F \leftarrow F \setminus Z$; // set of columns to be optimized

19 $p_{jt} = \begin{cases} 1 & \text{if } j = \arg \max_j \{w_{jt}(1 - \mathbf{P}_F)_{jt}, \forall t \in F\}, \\ p_{jt} & \text{otherwise} \end{cases}$; // updating

Algorithm 3: CSSLS Algorithm

Input : $\mathbf{B} \in \mathbb{R}^{J \times J}$, $\mathbf{C} \in \mathbb{R}^{J \times K}$, $\mathbf{P} \in \mathbb{R}^{J \times K}$

Output: \mathbf{X}

- 1 $M = \{1, \dots, K\}$, $N = \{1, \dots, J\}$, $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_K]$;
 - 2 Find the set of L unique columns in \mathbf{P} : $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_L] = \text{unique}\{\mathbf{P}\}$;
 - 3 $\mathbf{d}_j = \{t \in M : \mathbf{p}_t = \mathbf{u}_j\}$; // columns with identical passive sets
 - 4 **for** $j = 1, \dots, L$ **do**
 - 5 $\lfloor \mathbf{X}_{\mathbf{u}_j, \mathbf{d}_j} = (\mathbf{B}_{\mathbf{u}_j, \mathbf{u}_j})^{-1} [\mathbf{C}]_{\mathbf{u}_j, \mathbf{d}_j}$
-

$$D(\mathbf{Y}|\mathbf{A}\mathbf{X}) = \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_F^2, \tag{13}$$

where λ is a regularization parameter. In consequence, we have the Regularized FC-NNLS algorithm. The aim of using regularization for the discussed problems is rather to trigger the character of iterations than to stabilize ill-posed problems since the matrix \mathbf{A} is not expected to be ill-conditioned.

3.3 Regularized NNLS-NMF Algorithm

The Regularized NNLS-NMF algorithm is given by Algorithm 4. Motivated by

Algorithm 4: Regularized NNLS-NMF Algorithm

Input : $\mathbf{Y} \in \mathbb{R}^{I \times T}$, J - lower rank, λ_0 - initial regularization parameter, τ - decay rate, $\bar{\lambda}$ - minimal value of regularization parameter,

Output: Factors \mathbf{A} and \mathbf{X}

- 1 Initialize (randomly) \mathbf{A} and \mathbf{X} ;
 - 2 **repeat**
 - 3 $k \leftarrow k + 1$;
 - 4 $\lambda = \bar{\lambda} + \lambda_0 \exp\{-\tau k\}$; // Regularization parameter schedule
 - 5 $\mathbf{X} \leftarrow \text{fcnnls}(\mathbf{A}, \mathbf{Y}, \lambda)$; // Update for \mathbf{X}
 - 6 $d_j^{(X)} = \sqrt{\sum_{i=1}^T x_{ij}^2}$, $\mathbf{X} \leftarrow \text{diag}\{(d_j^{(X)})^{-1}\} \mathbf{X}$, $\mathbf{A} \leftarrow \mathbf{A} \text{diag}\{d_j^{(X)}\}$;
 - 7 $\bar{\mathbf{A}} \leftarrow \text{fcnnls}(\mathbf{X}^T, \mathbf{Y}^T, \lambda)$, $\mathbf{A} = \bar{\mathbf{A}}^T$; // Update for \mathbf{A}
 - 8 $d_j^{(A)} = \sqrt{\sum_{i=1}^I a_{ij}^2}$, $\mathbf{X} \leftarrow \text{diag}\{d_j^{(A)}\} \mathbf{X}$, $\mathbf{A} \leftarrow \mathbf{A} \text{diag}\{(d_j^{(A)})^{-1}\}$;
 - 9 **until** Stop criterion is satisfied;
-

[26], we propose to gradually decrease the regularization parameter λ with alternating iterations for NMF, starting from a large value λ_0 . Considering updates for \mathbf{X} and \mathbf{A} in terms of Singular Value Decomposition (SVD) of these matrices, it is obvious that if λ is large, only the right singular vectors that correspond to the largest singular values take part in the updating process. When the number of iterations is

large (in practice about 100), $\lambda \rightarrow \bar{\lambda}$, where we take $\bar{\lambda} = 10^{-12}$, and the singular vectors corresponding to the smallest singular values participate in the updates.

4 Experiments

The experiments are carried out for two databases of text documents. The first database (D1) is created from the *Reuters* database by randomly selecting 500 documents from the following topics: *acq*, *coffee*, *crude*, *eran*, *gold*, *interest*, *money-fx*, *ship*, *sugar*, *trade*. The database D1 has 5960 distinct words; thus $\mathbf{Y} \in \mathbb{R}^{5960 \times 500}$ and $J = 10$. The other database (D2) comes from the TopicPlanet document collection. We selected 178 documents classified into 6 topics: *air-travel*, *broadband*, *cruises*, *domain-names*, *investments*, *technologies*, which gives 8054 words. Thus $\mathbf{Y} \in \mathbb{R}^{8054 \times 178}$ and $J = 6$.

For clustering, we used the following algorithms: standard Lee-Seung NMF for the Euclidean distance [18], GD-CLS [22], Projected ALS [2, 7], Uni-orth NMF [10] separately for the matrix \mathbf{A} and \mathbf{X} , Bi-orth NMF [10], Convex-NMF [8], Newton NMF [7], k-means (from Matlab2008) with a cosine measure, LSI, and Regularized FC-NNLS (RFC-NNLS). All the algorithms are terminated after 20 iterations. For the RFC-NNLS algorithm, we set the following parameter: $\bar{\lambda} = 10^{-12}$, $\lambda_0 = 10^8$ and $\tau = 0.2$. All the algorithms are intensively tested under the software developed by M. Jankowiak [14].

Table 1 Mean-accuracy, standard deviations, and averaged elapsed time (in seconds) over 10 MC runs of the tested NMF algorithms.

Algorithm	TopicPlanet			Reuters		
	Accuracy	Std.	Time [sec]	Accuracy	Std.	Time [sec]
Lee-Seung NMF	0.8287	0.0408	0.09	0.6084	0.0542	0.28
GD-CLS	0.8528	0.0526	0.103	0.5982	0.0576	0.29
Projected ALS	0.8831	0.0260	0.105	0.6274	0.0333	0.3
Uni-orth (A)	0.8152	0.0484	40	0.4890	0.0263	23.9
Uni-orth (X)	0.7556	0.0529	3.62	0.5692	0.0528	7.3
Bi-orth	0.6315	0.0933	44.6	0.3896	0.0570	30.5
Convex-NMF	0.7680	0.0995	9.5	0.4624	0.0641	129.2
Newton-NMF	0.8691	0.0297	0.13	0.6446	0.0438	0.33
k-means	0.8551	0.0388	72.63	0.5720	0.029	214
LSI	0.8146	0	0.15	0.5994	0	0.39
RFC-NNLS	0.9326	0	0.17	0.8332	0.0018	0.35

Each tested algorithm (except for LSI) is run for 10 Monte Carlo (MC) trials with a random initialization. The algorithms are evaluated (Table 1) in terms of

the averaged accuracy of clustering, standard deviations, and the averaged elapsed time over 10 trials. The accuracy is expressed with a ratio of correctly classified documents to the total number of documents.

5 Conclusions

We proposed the Tikhonov regularized version of the FC-NNLS (RFC-NNLS) algorithm and efficiently applied it to text document clustering. The experiments demonstrate that the RFC-NNLS algorithm outperforms all the tested algorithms in terms of the accuracy and resistance to initialization. The computational complexity of the RFC-NNLS is comparable to fast algorithms such as the standard Lee-Seung NMF, ALS, Newton-NMF, and LSI.

Acknowledgment. This work was supported by the habilitation grant N N515 603139 (2010-2012) from the Ministry of Science and Higher Education, Poland.

References

- [1] Benthem, M.H.V., Keenan, M.R.: *J. Chemometr.* 18, 441–450 (2004)
- [2] Berry, M., Browne, M., Langville, A., Pauca, P., Plemmons, R.: *Comput. Stat. Data An.* 52, 155–173 (2007)
- [3] Bro, R., Jong, S.D.: *J. Chemometr.* 11, 393–401 (1997)
- [4] Buciu, I., Pitas, I.: Application of non-negative and local nonnegative matrix factorization to facial expression recognition. In: *Proc. Intl. Conf. Pattern Recognition (ICPR)*, pp. 288–291 (2004)
- [5] Cai, D., He, X., Wu, X., Bao, H., Han, J.: Locality preserving nonnegative matrix factorization. In: *Proc. IJCAI 2009*, pp. 1010–1015 (2009)
- [6] Cai, D., He, X., Wu, X., Han, J.: Nonnegative matrix factorization on manifold. In: *Proc. 8th IEEE Intl. Conf. Data Mining (ICDM)*, pp. 63–72 (2008)
- [7] Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.I.: *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley and Sons, Chichester (2009)
- [8] Ding, C., Li, T., Jordan, M.I.: *IEEE T. Pattern. Anal.* 32, 45–55 (2010)
- [9] Ding, C., Li, T., Peng, W.: Nonnegative matrix factorization and probabilistic latent semantic indexing: Equivalence, chi-square statistic, and a hybrid method. In: *Proc. AAAI National Conf. Artificial Intelligence (AAAI 2006)* (2006)
- [10] Ding, C., Li, T., Peng, W., Park, H.: Orthogonal nonnegative matrix tri-factorizations for clustering. In: *Proc 12th ACM SIGKDD Intl. Conf. Knowledge Discovery and Data Mining*, pp. 126–135. ACM Press, New York (2006)
- [11] Du, Q., Kopriva, I.: *Neurocomputing* 72, 2682–2692 (2009)
- [12] Heiler, M., Schnoerr, C.: *J. Mach. Learn. Res.* 7, 1385–1407 (2006)
- [13] Jain, A.K., Murty, M.N., Flynn, P.J.: *ACM Comput. Surv.* 31, 264–323 (1999)
- [14] Jankowiak, M.: Application of nonnegative matrix factorization for text document classification. MSc thesis (supervised by Dr. R. Zdunek), Wrocław University of Technology, Poland (2010) (in Polish)
- [15] Kim, H., Park, H.: *Bioinformatics* 23, 1495–1502 (2007)

- [16] Kim, H., Park, H.: *SIAM J. Matrix Anal. A* 30, 713–730 (2008)
- [17] Lawson, C.L., Hanson, R.J.: *Solving Least Squares Problems*. Prentice-Hall, Englewood Cliffs (1974)
- [18] Lee, D.D., Seung, H.S.: *Nature* 401, 788–791 (1999)
- [19] Li, T., Ding, C.: The relationships among various nonnegative matrix factorization methods for clustering. In: *Proc. 6th Intl. Conf. Data Mining (ICDM 2006)*, pp. 362–371. IEEE Computer Society, Washington DC, USA (2006)
- [20] O’Grady, P., Pearlmutter, B.: *Neurocomputing* 72, 88–101 (2008)
- [21] Sajda, P., Du, S., Brown, T.R., Stoyanova, R., Shungu, D.C., Mao, X., Parra, L.C.: *IEEE T. Med. Imaging* 23, 1453–1465 (2004)
- [22] Shahnaz, F., Berry, M., Pauca, P., Plemmons, R.: *Inform. Process. Manag.* 42, 373–386 (2006)
- [23] Sra, S., Dhillon, I.S.: *Nonnegative matrix approximation: Algorithms and Applications*. UTCS Technical Report TR-06-27, Austin, USA (2006), <http://www.cs.utexas.edu/ftp/pub/techreports/tr06-27.pdf>
- [24] Xu, W., Liu, X., Gong, Y.: Document clustering based on non-negative matrix factorization. In: *SIGIR 2003: Proc 26th Annual Intl ACM SIGIR Conf. Research and Development in Informaion Retrieval*, pp. 267–273. ACM Press, New York (2003)
- [25] Zdunek, R., Cichocki, A.: *Comput. Intel. Neurosci.* (939567) (2008)
- [26] Zdunek, R., Phan, A.H., Cichocki, A.: *Aust. J. Intel. Inform. Process. Syst.* 12, 16–22 (2010)

Distant Co-occurrence Language Model for ASR in Loose Word Order Languages

Jerzy Sas and Andrzej Zolnierek

Abstract. In the paper the problem of language modeling for automatic speech recognition in loose word order languages is considered. In loose word order languages classical n-gram language models are less effective, because the ordered word sequences encountered in the language corpus used to build the language models are less specific than in the case of strict word order languages. Because a word set appearing in the phrase is likely to appear in other permutation, all permutations of word sequences encountered in the corpus should be given additional likelihood in the language model. We propose the method of n-gram language model construction which assigns additional probability to word tuples being permutations of word sequences found in the training corpus. The paradigm of backoff bigram language model is adapted. The modification of typical model construction method consists in increasing the backed-off probability of bigrams that never appeared in the corpus but which elements appeared in the same phrases separated by other words. The proposed modification can be applied to any method of language model construction that is based on ML probability discounting. The performances of various LM creation methods adapted with the proposed way were compared in the application to Polish speech recognition.

1 Introduction

Speech recognition is one of the most natural and convenient methods of interaction of a human with computers. In order to achieve the accepted level of user

Jerzy Sas

Wroclaw University of Technology, Institute of Informatics,
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland

e-mail: jerzy.sas@pwr.wroc.pl

Andrzej Zolnierek

Wroclaw University of Technology, Faculty of Electronics,
Department of Systems and Computer Networks,
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland

e-mail: andrzej.zolnierek@pwr.wroc.pl

satisfaction with spoken man-machine dialog, high accuracy level of *Automatic Speech Recognition* (ASR) is desired. Experiences with practical implementation of ASR systems indicate that the *Word Error Rate* (WER) less than 5-10% must be reached in order the ASR-based software to be widely accepted [3]. While the accuracy of commercial and experimental ASR systems for English reaches 97-98%, the accuracy achieved typically for other languages is significantly lower. The main reason of difficulties in achieving low WER is the rich inflection of the language and loose word order permitted by the language syntax. Slavonic languages are particularly difficult for ASR due to these reasons [5], [6]. Because of rich inflection the language dictionary contains much more word forms than in the case of inflectionally-simple languages. Experiments described in [7] show that in order to obtain similar level of corpus coverage the dictionary for Russian have to contain almost 7 times more words than is needed for English. Firstly, big number of words in the dictionary leads to computational problems in the recognition process. Additionally phonetic differences of word form pronunciations are often insignificant what leads to problems in distinguishing word forms based on acoustic evidence. Another source of difficulties in ASR in languages like Polish is its loose word order. In Polish the word order is not so strictly imposed on the utterances by the language syntax as e.g. in English. If a sequence of words constitutes syntactically and semantically correct utterance then it is very likely that the permutation of these words also constitutes syntactically correct phrase. In result the language model perplexity of Polish is much higher than of English ([5], [10]-section 4.4). In result, typical *Language Models* (LM) based in counting n -grams appearing in the language corpus and estimating the probability of the next n -th word conditioned on the preceding sequence of $n - 1$ words are less effective in supporting ASR. This is because the actual conditional probability $p(w_i | w_{i-n+1}, \dots, w_{i-1})$ is more uniformly distributed among words w_i . In loose word order languages, instead of merely relying only on the sequences of words actually observed in the training corpus it seems to be reasonable to increase the n -gram probabilities of all word pairs that co-occur the utterances in the language corpus. In this way longer context of words can be taken into account in the language model.

Incorporation of distant context into the language model has been considered in a number of publications. One of possibilities is to take higher order n -gram models. Experiments described in [8] show that increasing n -gram order up to 6 improves the perplexity. It however requires very huge language corpus in order higher order n -gram probabilities to be estimated reliably. Another approach utilizes the observation that words once observed in the text are likely to be repeated again. It leads to the concept of dynamic language model where the probabilities of words just observed in the text are temporarily boosted ([1]).

The ASR accuracy can be also improved by applying a multistage approach, where the earlier stage provides the set of alternative word sequences and the subsequent stages re-evaluate the candidate sequence scoring and applying the longer distance word co-occurrence properties. One of possible approaches presented in [2] exploits the concept of semantic similarities of words. Such sequences likelihood is boosted which consist of word combinations that are semantically similar each

to other. The semantic similarity can be defined in various way but one of possibilities is to base it on word co-occurrence frequency in the language corpus. The concept presented in [2] was originally proposed for handwriting recognition but can be easily adapted to ASR needs.

The approach presented in this paper consists in modification of typical n-gram LM so as to boost the probabilities of n-grams consisting of words that co-occur in the corpus in utterances but are not direct neighbors. The modification can be applied to any backoff LM that is based on discounting. In typical approach ([4]) the probability mass obtained by discounting of n-gram probabilities estimated is distributed among all not observed n-grams proportionally to $(n - 1)$ -gram probabilities. In the approach proposed here more probability is allocated to co-occurring n-grams at the expense of lowering the probability allocated to n-grams which components did not occur in the same utterance in the language corpus. The factor by which the co-occurring n -gram probabilities are boosted is set so as to minimize the cross-perplexity computed using the subset of the language corpus excluded from the set used for LM building. Application of LM prepared in this way in the context of ASR reduces the risk of erroneous recognition of the words that are specific for a domain different than this one from which the actual utterance comes from. In this way the overall WER of speech recognition can be reduced.

The organization of the paper is as follows. In the section 2 the approach to ASR based on LM is shortly described. The next section presents selected LM smoothing techniques that will be later used as the basis for extensions with the approach presented here. In the section 4 the method of co-occurring n -gram probabilities boosting is presented in detail. Experimental evaluation and comparison of various LM models in Polish speech recognition is described in section 5. The last section presents conclusions and further research directions.

2 Automatic Speech Recognition with Acoustic and Language Models

Typical approach to the problem of automatic speech recognition consists in building acoustic models and language models combined into compound *hidden Markov model* (HMM), which can be considered as three-level system. On the lowest level, simple Markov models for individual Polish phonemes are created and trained. Uniform HMM topology for each phoneme is assumed. It consists of five states including initial and terminal ones. The state transition probabilities as well as the parameters of observations emission probability density functions for all phoneme HMMs are estimated using Baum-Welch procedure and the set of correctly transcribed training utterances. On the middle level the models of words are created by concatenating models of subsequent phonemes appearing in the phonetic translation of the word. Because the phoneme HMMs can be multiply applied in various words, training of HMM for the language consisting of a set of words does not require all words from the language to be presented during training. Then for each admissible word from the dictionary $\mathcal{D} = (w_1, w_2, \dots, w_N)$ we deal with the word

HMM which is built by concatenating HMMs for subsequent phonemes. Finally, on the highest level, the compound HMM of the whole utterance is built by connecting word HMMs in one language HMM. The probabilities of transition from the terminal state of a word HMM to initial state of another word HMM are taken from domain-specific n -gram language model and details of this procedure are presented in the next section. In automatic speech recognition we start with acquisition of the speech acoustic signal from the sound device and segment it into fragments separated by silence. In practice every utterance being recognized is converted in the sequence of vectors of observations (o_1, o_2, \dots, o_t) . Then finally the recognition with compound HMM consists in finding such a word sequence W^* which maximizes its conditional probability given the sequence of observations:

$$W^* = \arg \max_{w_{i_1}, w_{i_2}, \dots, w_{i_k} \in \mathcal{D}^+} P(w_{i_1}, w_{i_2}, \dots, w_{i_k} | o_1, o_2, \dots, o_t) \quad (1)$$

where \mathcal{D}^+ denotes the set of all nonempty sequences of words from the dictionary \mathcal{D} .

3 Backoff LM Smoothing

We will be considering here the languages being sets of sequences of words coming from the finite dictionary \mathcal{D} . Stochastic n -gram language model is the set of data that makes possible to estimate the probability of n -th word appearance provided that the sequence of preceding $n - 1$ words is known. In other words, LM provides the method to compute the estimation of:

$$p(w_i | w_{i-n+1}, w_{i-n+2}, \dots, w_{i-1}). \quad (2)$$

The sequence of n consecutive words is called n -gram. The conditional probabilities can be given explicitly in LM or they can be defined procedurally. The language model is usually constructed from the language corpus which is the sufficiently large set of exemplary phrases in the language being modeled. The most obvious way to find out the probability estimations is to count the occurrences of n -grams in the model and to apply *Maximal Likelihood* (ML) estimation:

$$p_{ML}(w_i | w_{i-n+1}, w_{i-n+2}, \dots, w_{i-1}) = \frac{c(w_{i-n+1}, w_{i-n+2}, \dots, w_{i-1}, w_i)}{c(w_{i-n+1}, w_{i-n+2}, \dots, w_{i-1})}, \quad (3)$$

where $c(w_1, w_2, \dots, w_n)$ is the number of n -gram w_1, w_2, \dots, w_n occurrences in the corpus. Due to limited size of the corpus, nonzero ML estimates can be obtained only for very limited set of n -grams. For this reason, in practice low n -gram orders n are used, in most cases n does not exceed 3. In our experiments we use bigram LM what corresponds to setting $n = 2$. Limiting the n -gram order still does not solve the data sparseness problem completely in the case of languages consisting of thousands of word. It is still very likely that many n -grams that may appear in typical language use are missing in the corpus at all or the number of their occurrences

is not big enough to allow for reliable ML estimation of related probabilities. To prevent underestimation of probabilities of missing n-grams the concept of *back-off* is applied. Backing off consists in using lower order n-gram probabilities in the case the number of occurrences of an n-gram in the corpus is not sufficient. In such case the probability (2) is approximated using the lower order n-gram probability $p(w_i|w_{i-n+2}, \dots, w_{i-1})$. In the bigram LM the probabilities $p(w_i|w_{i-1})$ are approximated by prior w_i word probabilities $p(w_i)$ which in most cases can be reliably estimated with ML estimators:

$$p(w_i) = \frac{c(w_i)}{\sum_{w_k \in \mathcal{D}} c(w_k)}. \quad (4)$$

While missing n-gram probabilities estimated with ML estimator are underestimated (nulled), the probabilities of n-grams occurring in the corpus only a few times are usually over-estimated. Therefore the concept of backoff is complemented with the concept of *discounting*. Probability discounting consists in subtracting some probability mass from probabilities (2) estimated with ML based on the formula (3). In result, for bigrams (4) occurring in the corpus, the probability $p_{ML}(w_i|w_{i-1})$ is replaced by the discounted probability $p_d(w_i|w_{i-1}) \leq p_{ML}(w_i|w_{i-1})$.

In discounted backoff LMs the probability mass discounted from the ML estimations of probabilities $p(w_i|w_{i-1})$ bigrams actually occurring in the corpus is distributed among words that never occurred as successors of w_{i-1} . The discounted probability mass $\beta(w)$ can be computed as:

$$\beta(w) = 1 - \sum_{w_k: c(w, w_k) > 0} p_d(w_k|w). \quad (5)$$

The conditional probabilities of words w_k that never appeared as successors of w are proportional to their prior probabilities computed according to (4). The probabilities $p(w_k|w)$ however have to sum up to 1.0 over all words w from the dictionary \mathcal{D} . Therefore the probabilities for bigrams (w, w_k) not observed in the corpus are finally computed as:

$$p(w_k|w) = \alpha(w)p(w_k), \quad (6)$$

where $\alpha(w)$ is calculated as:

$$\alpha(w) = \frac{\beta(w)}{\sum_{w_i: c(w, w_i) = 0} p(w_i)} = \frac{\beta(w)}{1 - \sum_{w_i: c(w, w_i) > 0} p(w_i)}. \quad (7)$$

Various schemes of discounting were proposed and tested in various LM applications ([4]). In our test for comparison purpose the following smoothing methods were chosen:

- Good-Turing estimate,
- absolute discounting,

¹ In the further part of the paper we will restrict our considerations to bigram language models.

- Kneser-Ney smoothing,
- modified Kneser-Ney smoothing.

The details of these methods can be found in ([4]) and their effectiveness in Polish ASR will be investigated in section 5. Let us briefly recall the ideas of above mentioned methods, using the notation from this paper ([4]).

The Good-Turing estimate states that for any bigram that occurs r times we should pretend that it occurs r^* times where

$$r^* = (r + 1) \frac{n_{r+1}}{n_r} \quad (8)$$

and where n_r is the number of bigrams that occur exactly r times in the training set. Consequently we can calculate the probability for a bi-gram δ with r counts

$$p_{GT}(\delta) = \frac{r^*}{N} \quad (9)$$

where $N = \sum_{r=0}^{\infty} n_r r^*$.

In **absolute discounting** the first-order distribution of Markov model is created by subtracting a fixed discount $0 \leq d \leq 1$ from each nonzero count, i.e.:

$$p_{ABS}(w_i | w_{i-1}) = \frac{\max[c(w_{i-1}, w_i) - d, 0]}{\sum_{w_i \in \mathcal{D}} c(w_{i-1}, w_i)} + (1 - \lambda_{w_{i-1}}) p_{ABS}(w_i). \quad (10)$$

To make this distribution sum to 1, we should take

$$1 - \lambda_{w_{i-1}} = \frac{d}{\sum_{w_i \in \mathcal{D}} c(w_{i-1}, w_i)} N_{1+}(w_{i-1}, w_i \bullet) \quad (11)$$

where the number of unique words that follow the history (w_{i-1}, w_i) is defined as

$$N_{1+}(w_{i-1} \bullet) = |\{w_i : c(w_{i-1}, w_i) > 0\}|. \quad (12)$$

The notation N_{1+} is meant to evoke the number of words that have one or more counts, and the \bullet is meant to evoke a free variable that is summed over.

Kneser-Ney smoothing is an extension of absolute discounting where the lower-order distribution that one combines with a higher-order distribution is built in another manner. For a bigram model we select a smoothed distribution p_{KN} that satisfies the following constraint on unigram marginals for all w_i

$$\sum_{w_{i-1} \in \mathcal{D}} p_{KN}(w_{i-1}, w_i) = \frac{c(w_i)}{\sum_{w_i \in \mathcal{D}} c(w_i)} \quad (13)$$

Kneser-Ney model can be presented in the same recursive way as ([10]) i.e.:

$$p_{KN}(w_i|w_{i-1}) = \frac{\max[c(w_{i-1}, w_i) - d, 0]}{\sum_{w_i \in \mathcal{D}} c(w_{i-1}, w_i)} + \kappa(w_{i-1}, w_i)p_{KN}(w_i). \quad (14)$$

where

$$\kappa(w_{i-1}, w_i) = \frac{d}{\sum_{w_i \in \mathcal{D}} c(w_{i-1}, w_i)} N_{1+}(w_{i-1}, w_i \bullet). \quad (15)$$

The unigram probabilities can be calculated as follows:

$$p_{KN}(w_i) = \frac{N_{1+}(\bullet w_i)}{N_{1+}(\bullet \bullet)} \quad (16)$$

where

$$N_{1+}(\bullet w_i) = | \{w_{i-1} : c(w_{i-1}, w_i) > 0\} | \quad (17)$$

and

$$N_{1+}(\bullet \bullet) = \sum_{w_i \in \mathcal{D}} (N_{1+}(\bullet w_i)). \quad (18)$$

In **modified Kneser-Ney smoothing**, the method proposed by Chen and Goodman in ([4]), instead of using a single discount d for all nonzero counts in KN smoothing we use three different parameters d_1 , d_2 and d_{3+} that are applied to bigrams with one, two, and three or more counts, respectively. Now the formula (14) turns into:

$$p_{MKN}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i) - d(c(w_{i-1}, w_i))}{\sum_{w_i \in \mathcal{D}} c(w_{i-1}, w_i)} + \gamma(w_{i-1})p_{MKN}(w_i) \quad (19)$$

where

$$d(c) = \begin{cases} 0 & \text{if } c = 0 \\ d_1 & \text{if } c = 1 \\ d_2 & \text{if } c = 2 \\ d_{3+} & \text{if } c \geq 3 \end{cases} \quad (20)$$

To make the distribution sum to 1, we take

$$\gamma(w_{i-1}) = \frac{d_1 N_1(w_{i-1} \bullet) + d_2 N_2(w_{i-1} \bullet) + d_{3+} N_{3+}(w_{i-1} \bullet)}{\sum_{w_i \in \mathcal{D}} c(w_{i-1}, w_i)} \quad (21)$$

where $N_2(w_{i-1} \bullet)$ and $N_{3+}(w_{i-1} \bullet)$ are defined analogously to $N_1(w_{i-1} \bullet)$ ([2]).

4 Probability Boosting for Indirectly Co-occurring n-Grams

The idea presented in this article consists in increasing the probability $p(w_k|w)$ for words w and w_k that occur in the corpus close each to other but do not necessarily appear in the adjacent positions. The idea behind this concept is motivated by the fact that in loose word order languages like Polish, if two words co-occur in the same utterance then it is likely that they will occur in other utterances on adjacent positions. So appearance of the co-occurrence of words in the corpus on

distant position can be an indication to increase the probability of the corresponding bigram. We assume here that the language corpus consists of clearly separated utterances. The probability of a bigram will be boosted if its two components co-occur in the same utterance.

Let us consider a single word $w \in \mathcal{D}$. Let $\mathcal{N}(w)$ denotes the set of words that appear at least once in the corpus directly after the word w , i.e. $\forall w_i \in \mathcal{N}(w) : c(w, w_i) > 0$. By $\mathcal{F}(w)$ we will denote the set of words that co-occur in at least one utterance with the word w but do not belong to $\mathcal{N}(w)$. $\mathcal{X}(w)$ denotes all remaining words from the dictionary ($\mathcal{X}(w) = \mathcal{D} \setminus \mathcal{N}(w) \setminus \mathcal{F}(w)$).

In the ordinary LM the probabilities for bigrams (w, w_k) for $w_k \in \mathcal{F}(w) \cup \mathcal{X}(w)$ are calculated according to (6), where $\alpha(w)$ is uniformly calculated using the formula (7). In order to boost the probability of bigrams consisting of words from $\mathcal{F}(w)$ their probabilities will be increased by the factor $\lambda > 1.0$, common for the whole model, i.e.:

$$\forall w_k \in \mathcal{F}(w) : p(w_k|w) = \lambda \alpha(w) p(w_k). \quad (22)$$

We assume that the total discounted probability mass $\beta(w)$ remains unchanged. To achieve this, the probabilities assigned to bigrams consisting of words from $\mathcal{X}(w)$ must be lowered appropriately. Now the probabilities (6) are multiplied by the factor $\bar{\lambda}(w) < 1.0$, which must be individually calculated for each w , so as to the total probability mass $\beta(w)$ is preserved:

$$\lambda \alpha(w) \sum_{w_k \in \mathcal{F}(w)} p(w_k) + \bar{\lambda}(w) \alpha(w) \sum_{w_k \in \mathcal{X}(w)} p(w_k) = 1 - \sum_{w_k \in \mathcal{N}(w)} p_d(w_k|w) = \beta(w). \quad (23)$$

Hence, $\bar{\lambda}(w)$ can be calculated as:

$$\bar{\lambda}(w) = \frac{1 - \sum_{w_k \in \mathcal{N}(w)} p_d(w_k|w) - \lambda \alpha(w) \sum_{w_k \in \mathcal{F}(w)} p(w_k)}{\alpha(w) \sum_{w_k \in \mathcal{X}(w)} p(w_k)} \quad (24)$$

Finally, the probability $p^*(w_k|w)$ that the modified language model assigns to a bigram (w, w_k) can be defined as:

$$p^*(w_k|w) = \begin{cases} p_d(w_k|w) & \text{if } w_k \in \mathcal{N}(w) \\ \lambda \alpha(w) p(w_k) & \text{if } w_k \in \mathcal{F}(w) \\ \bar{\lambda}(w) \alpha(w) p(w_k) & \text{if } w_k \in \mathcal{X}(w) \end{cases} \quad (25)$$

The value of λ , common for the whole model, can be computed so as to maximize the probability that the model assigns to the separated fragment of the language corpus. In order to do it the the corpus is divided into two parts: training part \mathcal{T} and evaluation part \mathcal{E} . The evaluation part is the set of word sequences s_1, s_2, \dots, s_n . Assume that each sequence starts with the specific start tag " $<s>$ " and ends with the end tag " $</s>$ ":

$$s_j = (<s>, w_1^{(j)}, w_2^{(j)}, \dots, w_{l(s_j)}^{(j)}, </s>). \quad (26)$$

The probability that LM assigns to the utterance s_j can be computed as:

$$P(s_j; LM(\mathcal{T}, \lambda)) = \prod_{k=1}^{l(s_j)+1} p^*(w_k^{(j)} | w_{k-1}^{(j)}; LM(\mathcal{T}, \lambda)), \quad (27)$$

where $w_0^{(j)} = \langle \mathbf{s} \rangle$ and $w_{l(s_j)+1}^{(j)} = \langle / \mathbf{s} \rangle$. The probability that the language the model $LM(\mathcal{T}, \lambda)$ assigns to the whole evaluation set \mathcal{E} is:

$$P(\mathcal{E}; LM(\mathcal{T}, \lambda)) = \prod_{s \in \mathcal{E}} P(s; LM(\mathcal{T}, \lambda)). \quad (28)$$

The value of λ is determined in an iterative procedure so as to maximize the probability (28).

5 Experimental Evaluation of LMs in Polish Speech Recognition

The aim of experiments carried out in the scope of this work was to evaluate the performance of LMs created using the method described in Section 4. The performance is assessed by a) the perplexity computed on the sentence set representative for a domain and b) by the word error rate of a speech recognizer which uses the LM being evaluated. The perplexity is the measure of LM quality. It is based on the average probability that the tested LM assigns to sentences in the test set computed "per word" (4). The lower is the LM perplexity, the better is the language stochastic properties approximation by LM. However, from the practical point of view the ultimate LM assessment should be rather determined by evaluating the LM contribution to the accuracy increase of the ASR process.

Four domains of Polish language which differ in complexity were used in the experiment:

- CT - texts from the domain of medical diagnostic image reporting; dictionary size - 23 thousands of words, corpus size - 22 MB,
- TL - texts related to the theory of Polish literature; dictionary size - 81 thousands of words, corpus size - 8 MB,
- GM - general medicine texts; dictionary size - 119 thousands of words, corpus size - 94 MB,
- PL - general purpose Polish language texts; dictionary size - 576 thousands of words, corpus size - 370 MB.

The experiment consisted of two stages. At the first stage, the performance of typical language models build with smoothing techniques described in 3 were compared by their perplexities and by WER of a speech recognizer based on the model being compared. Then the best smoothing method, taking into account WER is selected for further experiments. At the second stage, chosen method is combined with the backoff technique proposed here. The speech recognition accuracy obtained using

the modified model is compared with the corresponding accuracy achieved with the conventional LM.

The acoustic model for ASR was created in speaker-dependent manner as a trigram model using the speech samples recorded by a single male speaker. The total duration of training utterances is about 5 hours. For ASR accuracy testing the individual set of utterances was used for each domain LM. The duration of the utterances in the test set was in the range 30-40 minutes. HTK toolkit ([13]) was applied to build the acoustic model. The open source recognition engine Julius ([12]) was used as the speech recognizer.

The results of the first stage of the experiment are shown in Tables 1 and 2. Just for the sake of comparison, results obtained with the unigram LM are also shown. No smoothing is applied in the case of unigram LM.

Table 1 The perplexity of models based on various smoothing methods

Discounting method	CT	TL	GM	PL
Unigram not smoothed	751.5	2688.3	1596.3	4233.6
Absolute	35.9	748.4	69.2	672.1
Good-Turing	35.9	733.2	68.2	665.4
Kneser-Ney	34.5	713.0	65.7	633.3
Kneser-Ney modified	34.9	720.4	66.6	633.5

Table 2 ASR accuracy obtained with models based on various smoothing methods

Discounting method	CT	TL	GM	PL
Unigram not smoothed	93.7	89.1	92.7	90.3
Absolute	95.3	90.6	94.2	91.6
Good-Turing	95.2	90.9	94.8	92.2
Kneser-Ney	95.9	91.6	95.1	92.6
Kneser-Ney modified	95.6	91.7	95.0	91.9

It can be observed that there are no significant differences between bigram models created with various discounting methods. Despite the loose word order of Polish language, bigram models have much lower perplexity than unigram models. The bigram/unigram perplexity ratios for Polish are similar to these ones reported for English in [10]. ASR in Polish is however much less sensitive to LM perplexity than as it is in the case of English. Authors in [4] claim that the increase of cross-entropy (which is a logarithm of perplexity) by 0.2 results in the absolute increase of WER by 1%. Our experiments (in particular - comparisons of perplexities and WERs for

CT and PL models) show that increase of cross-perplexity by 4.2 results in the increase of WER by only about 3.5%. Also the superiority of modified Kneser-Ney smoothing method over all other methods compared here reported in [4] is not confirmed in the case of Polish language. The performance on modified Kneser-Ney model is slightly worse than the performance of the original Kneser-Ney model. Although the performances of all bigram models are similar, the original Kneser-Ney model achieves best results both in perplexity and WER comparisons. This one was used for further experiments.

At the second stage the Kneser-Ney smoothing was combined with backoff technique described in Section 4. The accuracy of speech recognizer was then evaluated for the modified model. ASR accuracies in various language domains are presented in Table 3.

Table 3 ASR accuracy obtained with the model based on indirectly co-occurring bigram probability boosting

Discounting method	CT	TL	GM	PL
Original Kneser-Ney	95.9	91.6	95.1	92.6
Indirectly co-occurring bigrams boosting	95.9	92.1	95.4	92.9
Relative WER reduction	0.0%	5.9%	6.1%	4.1%

The application of indirectly co-occurring bigrams boosting resulted in small but observable improvement of ASR accuracy in the case of all language domains except of the simplest CT model. The average relative WER reduction is about 4%.

6 Conclusions, Further Works

In the paper new method of language modeling for automatic speech recognition in loose word order languages is proposed. It consists in increasing the backed-off probability of bigrams that never appeared in the corpus but which elements appeared in the same phrases separated by other words. The proposed modification can be applied to any method of language model construction that is based on ML probability discounting. The performances of various LM creation methods adapted with the proposed method were compared in the application to Polish speech recognition. The results of WER are promising, i.e. the average WER is reduced about 4%, but of course further empirical studies, for different domains of Polish language are needed. Another parameter which has the influence for the results is the size of the context window and it also requires further tests.

References

- [1] Jelinek, F., Merialdo, B., Roukos, S., Strauss, M.: A dynamic language model for speech recognition. In: *Proceedings of the Workshop on Speech and Natural Language, HLT 1991*, Association for Computational Linguistics, pp. 293–295 (1991)
- [2] Piasecki, M., Broda, B.: Correction of medical handwriting OCR based on semantic similarity. In: Yin, H., Tino, P., Corchado, E., Byrne, W., Yao, X. (eds.) *IDEAL 2007*. LNCS, vol. 4881, pp. 437–446. Springer, Heidelberg (2007)
- [3] Devine, E.G., Gaehde, S.A., Curtis, A.C.: Comparative Evaluation of Three Continuous Speech Recognition Software Packages in the Generation of Medical Reports. *Journal of American Medical Informatics Association* 7(5), 462–468 (2000)
- [4] Chen, S.F., Goodman, S.: An empirical study of smoothing techniques for language modeling. *Computer Speech and Language* (13), 359–394 (1999)
- [5] Ziolkowski, B., Skurzok, D., Ziolkowski, M.: Word n-grams for Polish. In: *Proc. of 10th IASTED Int. Conf. on Artificial Intelligence and Applications (AIA 2010)*, pp. 197–201 (2010)
- [6] Mauces, M., Rotownik, T., Zemljak, M.: Modelling Highly Inflected Slovenian Language. *International Journal of Speech technology* 6, 254–257 (2003)
- [7] Whittaker, E.W.D., Woodland, P.C.: Language modelling for Russian and English using words and classes. In: *Computer Speech and Language*, vol. 17, pp. 87–104. Elsevier Academic Press, Amsterdam (2003)
- [8] Joshua, T., Goodman, J.T.: A Bit of Progress in Language Modeling Extended Version. Machine Learning and Applied Statistics Group Microsoft Research. Technical Report, MSR-TR-2001-72 (2001)
- [9] Katz, S.: Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer. *IEEE Transactions On Acoustics, Speech and Signal Processing* ASP-35(3), 400–401 (1987)
- [10] Jurafsky, D., Matrin, J.: *Speech and language processing. An introduction to natural language processing*. In: *Computational Linguistics and Speech Recognition*, Pearson Prentice Hall, New Jersey (2009)
- [11] Gale, A., Sampson, G.: Good-Turing frequency estimation without tears. *Journal of Quantitative Linguistics* 2, 217–239 (1995)
- [12] Lee, A., Kawahara, T., Shikano, K.: Julius - an Open Source Real-Time Large Vocabulary Recognition Engine. In: *Proc. of European Conference on Speech Communication and Technology (EUROSPEECH)*, pp. 1691–1694 (2001)
- [13] Young, S., Everman, G.: *The HTK Book (for HTK Version 3.4)*. Cambridge University Engineering Department (2009)

Author Index

- Andrysiak, Tomasz 727
Anne, K.R. 685
Antosik, Bartłomiej 197
Armano, Giuliano 167
Arnon, Parunyu 747
Augustyniak, Piotr 529, 539
- Bakała, Marcin 441
Beheshti Shirazi, Ali Asghar 657
Bhatt, Mehul 471
Bielecka, Marzena 479
Bielecki, Andrzej 479
Bogavarapu, L.N.P. 685
Brodowska, Magdalena 517
Burduk, Robert 219
- Chmielnicki, Wiesław 499
Cholewa, Michał 307
Choraś, Michał 61, 727
Chudzian, Paweł 99
Cislariu, Mihaela 449
Cudek, Paweł 605
- Doros, Marek 549
Doroz, Rafał 3
Droege, Detlev 375
Dziękowski, Bartłomiej 249
- Fabijańska, Anna 73, 441
Feinen, Christian 375
- Fevens, Thomas 623
Filipczuk, Paweł 613
Fischer, Matthias 317
Fischer, Christoph 317
Flasiński, Mariusz 187
Flisowska-Wiercik, Barbara 697
Florek, Andrzej 647
Forczmański, Paweł 327
Frejlichowski, Dariusz 327
- Gajek, Szymon 471
García, I. 259, 347
Głomb, Przemysław 307
Głomb, Przemysław 365
Gordan, Mihaela 449
Goszczyńska, Hanna 549
Grabska, Ewa 471
Grąbczewski, Krzysztof 147
Grzegorek, Maria 239
Grzegorek, Marcin 375
Grzymala-Busse, Jerzy W. 605
Guzmán, E. 259, 347
- Habibzadeh, Mehdi 623
Hatami, Nima 167
Henzel, N. 559
Hippe, Zdzisław S. 605
Hoser, Paweł 549
Hudyma, Elżbieta 249

- Iwanowski, Marcin 385, 411
- Jabłoński, Grzegorz 127
Jeleń, Łukasz 623
Jezewski, Michał 177
Jurek, Janusz 187
- Kabaciński, Rafał 51
Kasprzak, Włodzimierz 737
Kawulok, Michał 269
Kawulok, Jolanta 269
Klukowski, Leszek 91
Kolakowska, Agata 667
Kolebska, Krystyna 549
Komorowski, Jacek 13
Korkosz, Mariusz 479
Kostka, Paweł S. 597
Kostorz, Iwona 3
Koszmider, Tomasz 441
Kowal, Marek 613
Kowalczyk, Leszek 549
Kowalski, Mateusz 51
Kozik, Rafał 61, 459
Kraft, Marek 431
Krawczyk, Bartosz 207
Krol, Maciej 647
Krzeszowski, Tomasz 297
Krzyżak, Adam 623
Kurzynski, Marek 197
Kwolek, Bogdan 297
- Łabędź, Piotr 507
Lazarek, Jagoda 357
Leski, Jacek 177
Łęski, J. 559
Lis, Robert A. 707
Luszczkiewicz-Piatek, Maria 337
- Małkiewicz, Agnieszka 569
Mendoza, M. 259
Michalak, Krzysztof 249
Michalak, Marcin 137, 395
Mika, Barbara T. 597
- Mokrzycki, W.S. 279
Momot, A. 559
Momot, M. 559
Obuchowicz, Andrzej 613
Okarma, Krzysztof 287
Ozimek, Agnieszka 507
- Palacz, Wojciech 471
Pamula, Wiesław 677
Paradowski, Mariusz 403
Paulus, Dietrich 375
Pogrebnyak, Oleksiy 347
Prevete, R. 157
Przybysz, Paweł 737
- Radomski, Dariusz 569
Raniszewski, Marcin 119
Renk, Rafał 727
Rokita, Przemysław 13
Romaszewski, Michał 365
Rutkowska, Iwona 697
Rybniček, Marlies 317
- Saganowski, Łukasz 727
Salca, Victor Eugen 449
Sas, Jerzy 767
Schaefer, Gerald 229
Schmidt, Adam 431
Seekaew, Piyawan 747
Sierszeń, Artur 109
Simić, Dragan 717
Simić, Svetlana 717
Skomorowski, Marek, 479
Śledzik, J. 559
Śledzik, Joanna 587
Ślusarczyk, Grażyna 83
Śluzek, Andrzej 403
Smacki, Łukasz 41
Smiatacz, Maciej 23
Smolka, Bogdan 269, 357
Sohrabi, Shahin 657
Sprężak, Kamila 479
Stanek, Michał 249

- Stapor, Katarzyna 499
Stelengowska, Mirosława 587
Strug, Barbara 489
Strzecha, Krzysztof 441
Sturgulewski, Łukasz 109
Swiercz, Michał 385
Świtoński, Adam 395
Szczepaniak, Piotr S. 357, 579
Szczepanski, Marek 421
Szczurek, Andrzej 697
Szecówka, Przemysław M. 697
- Tatol, M. 279
Tessitore, G. 157
Tkacz, Ewaryst J. 597
Tomczyk, Arkadiusz 579
- Vaddi, R.S. 685
Vankayalapati, H.D. 685
Vlaicu, Aurel 449
- Wieclaw, Łukasz 33
Wojciechowski, Konrad 297
Wojciechowski, Wadim 479
Wongsirichot, Thakerng 747
Wozniak, Michał 207
Wrobel, Krzysztof 41
- Zambrano, J.G. 347
Zdunek, Rafał 757
Zieliński, Bartosz 479
Zielinski, Bartłomiej 411
Zolnierek, Andrzej 767