Advances in Intelligent Systems and Computing 313

Ryszard S. Choraś *Editor*

Image Processing & Communications Challenges 6

Springer

Advances in Intelligent Systems and Computing

Volume 313

Series editor

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

About this Series

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

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

Advisory Board

Chairman

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

Members

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

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

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

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

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

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

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

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

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

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

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

More information about this series at http://www.springer.com/series/11156

Ryszard S. Choraś
Editor

Image Processing & Communications Challenges 6

Springer

Editor
Ryszard S. Choraś
Institute of Telecommunications
University of Technology & Life Sciences
Bydgoszcz
Poland

ISSN 2194-5357 ISSN 2194-5365 (electronic)
ISBN 978-3-319-10661-8 ISBN 978-3-319-10662-5 (eBook)
DOI 10.1007/978-3-319-10662-5

Library of Congress Control Number: 2014948175

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

Preface

It is with great pleasure that I welcome you all to the proceedings of the 6th International Image Processing and Communications Conference (IP&C 2014) held in Bydgoszcz, 10–12 September 2014.

At the same time is growing interest in Image Processing and Communications and its applications in science, engineering, medical image analysis and biometrics.

IP&C is a central topic in contemporary computer and IT science and are interdisciplinary areas, where researchers and practitioners may find interesting problems which can be successfully solved.

The present book includes a set of papers on actual problems of research and application of image processing and communications technologies. I sincerely hope that this book provides a good view into research presented at the IP&C 2014 Conference and will offer opportunities for professional growth.

The book consists of two sections:

1. Image processing
2. Communications.

Finally, I would like to thank all authors, reviewers and participants, who contributed to the high quality of the IP&C Conference and scientific exchanges.

Bydgoszcz
September 2014 *Ryszard S. Choraś*

Table of Contents

Part I: Image Processing

Two-Dimensional Hidden Markov Models in Road Signs Recognition 3
Janusz Bobulski

Evaluating the Mutual Position of Objects on the Visual Scene Using Morphological Processing and Reasoning 13
Arkadiusz Cacko, Marcin Iwanowski

Vascular Biometry 21
Ryszard S. Choraś

Clustering-Based Retrieval of Similar Outfits Based on Clothes Visual Characteristics 29
Piotr Czapiewski, Paweł Forczmański, Dariusz Frejlichowski, Radosław Hofman

Improving Shape Retrieval and Classification Rates through Low-Dimensional Features Fusion 37
Paweł Forczmański

Accelerating the 3D Random Walker Image Segmentation Algorithm by Image Graph Reduction and GPU Computing 45
Jarosław Gocławski, Tomasz Węgliński, Anna Fabijańska

Computed Tomography Images Denoising with Markov Random Field Model Parametrized by Prewitt Mask 53
Michał Knas, Robert Cierniak

Neural Video Compression Algorithm 59
Michał Knop, Piotr Dobosz

Discovering Important Regions of Cytological Slides Using Classification Tree 67
Marek Kowal, Andrzej Marciniak, Roman Monczak, Andrzej Obuchowicz

Gaussian Mixture Model Based Non-local Means Technique for Mixed Noise Suppression in Color Images 75
Maria Luszczkiewicz-Piatek

Robust Image Retrieval Based on Mixture Modeling of Weighted Spatio-color Information 85
Maria Luszczkiewicz-Piatek, Bogdan Smolka

Automatic Classification of Tahitian Pearls 95
Martin Loesdau, Sébastien Chabrier, Alban Gabillon

Noise Reduction in Ultrasound Images Based on the Concept of Local Neighborhood Exploration 103
Krystyna Malik, Bernadetta Machala, Bogdan Smołka

Viterbi Algorithm for Noise Line Following Robots 111
Przemysław Mazurek

Influence of Impulse Noise on ST–TBD and Viterbi Track–Before–Detect Algorithms 119
Przemysław Mazurek

Hybrid Shape Descriptors for an Improved Weld Defect Retrieval in Radiographic Testing 127
Nafaa Nacereddine, Djemel Ziou

On the Usefulness of Combined Metrics for 3D Image Quality Assessment 137
Krzysztof Okarma

Application of Super-Resolution Algorithms for the Navigation of Autonomous Mobile Robots 145
Krzysztof Okarma, Mateusz Tecław, Piotr Lech

Image Processing with Process Migration 153
Zdzisław Onderka, Dariusz Półchłopek

Automatic Extraction of Business Logic from Digital Documents 161
Paweł Pieniążek

Comparison of Assessment Regularity Methods Dedicated to Isotropic Cells Structures Analysis 169
Adam Piórkowski, Przemysław Mazurek, Jolanta Gronkowska–Serafin

Towards Automated Cell Segmentation in Corneal Endothelium Images 179
Adam Piórkowski, Jolanta Gronkowska-Serafin

A Texture-Based Energy for Active Contour Image Segmentation 187
Daniel Reska, Cezary Boldak, Marek Kretowski

Object Localization and Detection Using Variance Filter 195
Grzegorz Sarwas, Sławomir Skoneczny

The Impact of the Image Feature Detector and Descriptor Choice on Visual SLAM Accuracy 203
Adam Schmidt, Marek Kraft

Face Recognition: Shape versus Texture 211
Maciej Smiatacz

Spatio-temporal Digital Path Approach to Video Enhancement 219
Marek Szczepański

Emotion Recognition for Affect Aware Video Games 227
Mariusz Szwoch, Wioleta Szwoch

Registration and Analysis of Data during Registration and Exposure Process 237
Jarosław Zdrojewski, Adam Marchewka, Rocío Pérez de Prado

Detection of Defects in Carbon-Fiber Composites Using Computer-Vision-Based Processing of Microwave Maps 245
Bartłomiej Zieliński, Marcin Iwanowski, Bartłomiej Salski, Szymon Reszewicz

Part II: Communications

Network Anomaly Detection Based on ARFIMA Model 255
Tomasz Andrysiak, Łukasz Saganowski

Is European Broadband Ready for Smart Grid? 263
Kartheepan Balachandran, Jens Myrup Pedersen

A Simulator Concept for Cloud Computing Infrastructure 269
Sławomir Hanczewski, Marta Kędzierska, Maciej Piechowiak

The New Model of Limited Availability Group 277
Sławomir Hanczewski, Wojciech Słowik

Hardening Web Applications against SQL Injection Attacks Using Anomaly Detection Approach 285
Rafał Kozik, Michał Choraś, Witold Hołubowicz

Versatile Remote Access Environment for Computer Networking Laboratory 293
Karol Kuczyński, Rafał Stęgierski, Waldemar Suszyński, Michael Pellerin

A Low Bitrate Video Transmission for IP Networks 301
Piotr Lech

Dynamic Routing and Spectrum Assignment for Varying Traffic in Flexible Optical Networks 309
Ireneusz Olszewski

Reduction of Reference Set for Network Data Analyzing Using the Bubble Algorithm 319
Artur Sierszeń

Author Index 329

Part I

Image Processing

Two-Dimensional Hidden Markov Models in Road Signs Recognition

Janusz Bobulski

Czestochowa University of Technology, Institute of Computer and Information Science, Dabrowskiego Street 73, 42-200 Czestochowa, Poland
januszb@icis.pcz.pl

Abstract. This paper presents an automatic road sign recognition system. The system bases on twodimensional hidden Markov models. First, a study of the existing road sign recognition research is presented. In this study, the issues associated with automatic road sign recognition are described, the existing methods developed to tackle the road sign recognition problem are reviewed, and a comparison of the features of these methods is given. Second, the created road sign recognition system is described. The system is able to recognize the road signs, which was detected earlier. The system makes use of two dimensional discrete wavelet transform for features extraction of road signs. In recognition process system bases on two dimensional hidden Markov models. The experimental results demonstrate that the system is able to achieving an average recognition rate of 83% using the two-dimensional hidden Markov models and the wavelet transform.

1 Introduction

Lots of research in the field of automatic road sign detection and recognition systems has been done for last thirty years. The main object of such systems is to warn them of presence of road signs in different ways, because drivers may not notice the presence of them. It is important to note to some of the common problems that are appear in road signs detection. The main problem to be overcome is caused by the variable lighting conditions of a scene in a natural environment and the possible rotation of the signs. An automatic system should be able to detect signs in different condition and position [6,15]. The first paper in the field of road sign recognition appears in Japan in 1984. Since then, great amount of methods have been developed to locate and identify road signs. A road sign is individual in its colour, shape, and appearance. Most of the existing methods use the combination of these properties to determine the meaning of road sign content [16]. Detection of road sign are mainly based on colour criteria [6] or shape information [20]. As a result, the colour space plays an lead role. The popular RGB color space are used in [3]. However, the RGB space is not optimized for problems of road signs recognition, because it is sensitive to lighting conditions changes [5]. Colour spaces that are more independent to such changes is HSV, and it is used in [13]. They applied a threshold over a hue,

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_1

saturation, value representation of the image to find regions with a road sign. After a road sign has been detected, it should be recognized from a large set of possible patterns using some of classification method. The symbol of detected sign must be suitably represented for a specific classification method that is usually done by extracting some features from the image. Examples of features extracting techniques are histograms [22] and wavelets [8]. After the choosing of appropriate features, many classification methods have been proposed, different kinds of neural networks [22,8,17], SVM [13] or hidden Markov model [7]. This paper presents an automatic road sign recognition system with following properties:(i) the system uses two dimensional wavelet transform of second level decomposition for features extraction, (ii) the classification module bases on two dimensional hidden Markov models, which work with two dimensional data.

Hidden Markov models (HMM) are widely apply in data classification. They are used in speech recognition, character recognition, biological sequence analysis, financial data processing, texture analysis, face recognition, etc. This widely application of HMM is result of its effectiveness. An extension of the HMM to work on two-dimensional data is 2D HMM. A 2D HMM can be regarded as a combination of one state matrix and one observation matrix, where transition between states take place according to a 2D Markovian probability and each observation is generated independently by the corresponding state at the same matrix position. It was noted that the complexity of estimating the parameters of a 2D HMMs or using them to perform maximum a posteriori classification is exponential in the size of data. Similar to 1D HMM, the most important thing for 2D HMMs is also to solve the three basic problems, namely, probability evolution, optimal state matrix and parameters estimation.

When we process one-dimensional data, we have good tools and solution for this. Unfortunately, this is unpractical in image processing, because the images are two-dimensional. When you convert an image from 2D to 1D , you lose some information. So, if we process two-dimensional data, we should apply twodimensional HMM, and this 2D HMM should works with 2D data. One of solutions is pseudo 2D HMM [2,23,11]. This model is extension of classic 1D HMM. There are super-states, which mask one-dimensional hidden Markov models (Fig. 1). Linear model is the topology of superstates, where only self transition and transition to the following superstate are possible. Inside the superstates there are linear 1D HMM. The state sequences in the rows are independent of the state sequences of neighboring rows. Additional, input data are divided to the vector. So, we have 1D model with 1D data in practise.

Other approach to image processing use twodimensional data present in works [12] and [9]. The solutions base on Markov Random Fields (MRF) give good results for classification and segmentation, but not in pattern recognition. Interesting results showed in paper [24]. This article presents analytic solution and proof of correctness two-dimensional HMM. But this 2D HMM is similar to MRF, works with onedimensional data and can be apply only for left-right type of HMM. This article presents real solution for 2D problem in HMM. There is show true 2D HMM which processes 2D data.

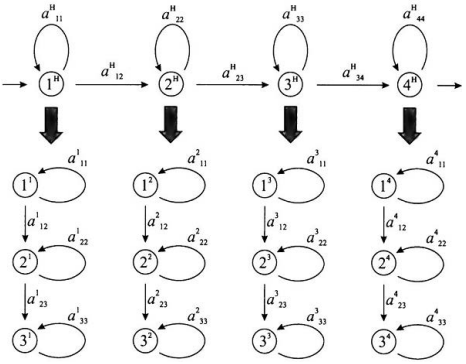


Fig. 1. Pseudo 2D HMM [2]

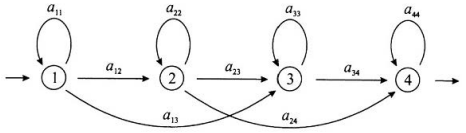


Fig. 2. One-dimensional HMM

2 Classic 1D HMM

HMM is a double stochastic process with underlying stochastic process that is not observable (hidden), but can be observed through another set of stochastic processes that produce a sequence of observation [18]. Let $O = \{O_1, \ldots, O_T\}$ be the sequence of observation of feature vectors, where T is the total number of feature vectors in the sequence. The statistical parameters of the model may be defined as follows [10]:

- The number of states of the model, N
- The number of symbols M
- The transition probabilities of the underlying Markov chain, $A = \{a_{ij}\}, 1 \leq i, j \leq N$, where a_{ij} is the probability of transition from state i to state j
- The observation probabilities, $B = \{b_{jm}\} 1 \leq j \leq N, 1 \leq m \leq M$ which represents the probability of gnerate the m_{th} symbol in the j_{th} state.
- The initial probability vector, $\Pi = \{\pi_i\} 1 \leq i \leq N$.

Hence, the HMM requires three probability measures to be defined, A, B, Π and the notation $\lambda = (A, B, \Pi)$ is often used to indicate the set of parameters of the model. In the proposed method, one model is made for each part of the face. The parameters of the model are generated at random at the beginning. Then they are estimated with Baum-Welch algorithm, which is based on the forward-backward algorithm. The forward algorithm calculates the coefficient $\alpha_t(i)$ (probability of observing the partial sequence $(o_1, , o_t)$ such that state q_t

is i). The backward algorithm calculates the coefficient $\beta_t(i)$ (probability of observing the partial sequence $(o_{t+1},,o_T)$ such that state q_t is i). The Baum-Welch algorithm, which computes the λ, can be described as follows [10]:

1. Let initial model be λ_0
2. Compute new λ based on λ_0 and observation O
3. If $\log(P(O|\lambda)) - \log(P(O)|\lambda_0) < DELTA$ stop
4. Else set $\lambda \rightarrow \lambda_0$ and go to step 2.

The parameters of new model λ, based on λ_0 and observation O, are estimated from equation of Baum-Welch algorithm [18], and then are recorded to the database.

2.1 Three Basic Problems

There are three fundamental problems of interest that must be solved for HMM to be useful in some applications. These problems are the following:

1. Given observation $O = (o_1, o_2,, o_T)$ and model $\lambda = (A, B, \Pi)$, efficiently compute $P(O|\lambda)$
2. Given observation $O = (o_1, o_2,, o_T)$ and model λ find the optimal state sequence $q = (q_1, q_2,, q_T)$
3. Given observation $O = (o_1, o_2,, o_T)$, estimate model parameters $\lambda = (A, B, \Pi)$ that maximize $P(O|\lambda)$

To create a system of pattern recognition is necessary to solve the problem 1 and 3.

2.2 Solution to Problem 1

Solution for problem 1 is well know forwardbackward algorithm [10].

Forward algorithm

– Define forward variable $\alpha_t(i)$ as:

$$\alpha_t(i) = P(o_1, o_2,, o_t, q_t = i|\lambda) \tag{1}$$

– $\alpha_t(i)$ is the probability of observing the partial sequence $(o_1, o_2,, o_t)$ such that the state q_t is i

Backward algorithm [10]

– Define backward variable $\beta_t(i)$ as:

$$\beta_t(i) = P(o_{t+1}, o_{t+2},, o_T, q_t = i|\lambda) \tag{2}$$

– $\beta_t(i)$ is the probability of observing the partial sequence $(o_1, o_2,, o_t)$ such that the state q_t is i

2.3 Solution to Problem 3

Baum-Welch Algorithm [10]:

– Define $\xi(i,j)$ as the probability of being in state i at time t and in state j at time $t+1$

$$\xi(i,j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{P(O|\lambda)} = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)} \tag{3}$$

– Define $\gamma(i)$ as the probability of being in state i at time t, given observation sequence.

$$\gamma_t(i) = \sum_{j=1}^{N}\xi_t(i,j) \tag{4}$$

– $\sum_{t=1}^{T}\gamma_t(i)$ is the expected number of times state i is visited
– $\sum_{t=1}^{T-1}\xi_t(i)$ is the expected number of transition from state i to j

Update rules:

– Ů $\bar{\pi}_i$ = expected frequency in state i at time $(t=1) = \gamma_1(i)$
– Ů $\bar{a}_{ij}$ = (expected number of transition from state i to state j)/(expected number of transitions from state i):

$$\hat{a}_{ij} = \frac{\sum_t \xi_t(i,j)}{\sum_t \gamma_t(i)} \tag{5}$$

– $\bar{b}_j(k)$ = (expected number of times in state j and oserving symbol k)/ (expected number of times in state j):

$$\bar{b}_j(k) = \frac{\sum_{t,o_t=k}\gamma_t(j)}{\sum_t \gamma_t(j)} \tag{6}$$

3 2D HMM

In paper [24], Yujian proposed definitions and proofs of 2D HMM. He has presented several analytic formulae for solving the three basic problems of 2-D HMM. Solution to Problem 2 is usefull, and Viterbi algorithm can be easily adopted to image recognition with two dimensional input data. Unfortunetly, solution to problem 1 and 3 may be use only with one dimensional data -observation vector. Besides presented solutions are for Markov model type "left-right", and not ergodic. So, we present solution to problems 1 and 3 for two dimensional data, which is sufficient to build a image recognition system. The statistical parameters of the 2D model (Fig. 3):

– The number of states of the model N^2.
– The number of data streams $k_1 \times k_2 = K$.

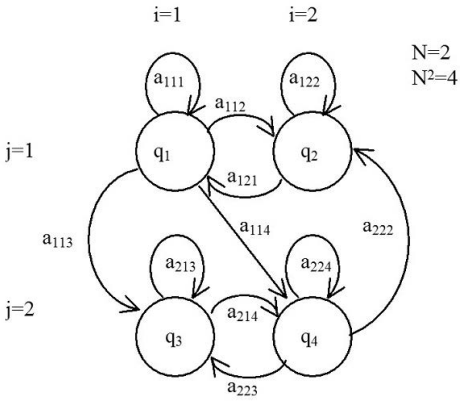


Fig. 3. Two-dimensional ergodic HMM

- The number of symbols M.
- The transition probabilities of the underlying Markov chain, $A = \{a_{ijl}\}, 1 \leq i, j \leq N$, $1 \leq l \leq N^2$, where a ij is the probability of transition from state ij to state l.
- The observation probabilities, $B = \{b_{ijm}\}, 1 \leq i, j \leq N, 1 \leq m \leq M$ which represents the probability of gnerate the m_{th} symbol in the ij_{th} state.
- The initial probability, $\Pi = \{\pi_{ijk}\}, 1 \leq i, j \leq N, 1 \leq k \leq K$.
- Oservation sequance $O = \{o_t\}, 1 \leq t \leq T, o_t$ is square matrix simply observation with size $k_1 \times k_2 = K$.

3.1 Solution to 2D Problem 1

Forward Algorithm

- Define forward variable $\alpha_t(i, j, k)$ as:

$$\alpha_t(i, j, k) = P(o_1, o_2, , 0_t, q_t = ij|\lambda) \tag{7}$$

- $\alpha_t(i, j, k)$ is the probability of observing the partial sequence $(o_1, o_2, , o_t)$ such that the state q_t is i, j for each k_{th} strem of data
- Induction
  1. Initialization:

$$\alpha_1(i, j, k) = \pi_{ijk} b_{ij}(o_1) \tag{8}$$

  2. Induction:

$$\alpha_{t+1}(i, j, k) = \left[\sum_{l=1}^{N} \alpha_t(i, j, k) a_{ijl}\right] b_{ij}(o_{t+1}) \tag{9}$$

  3. Termination:

$$P(O|\lambda) = \sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_T(i, j, k) \tag{10}$$

3.2 Solution to 2D Problem 3

Parameters reestimation Algorithm:

– Define $\xi(i,j,l)$ as the probability of being in state ij at time t and in state l at time $t+1$ for each k_{th} strem of data

$$\begin{aligned}\xi_t(i,j,l) &= \frac{\alpha_t(i,j,k)a_{ijl}b_{ij}(o_{t+1})\beta_{t+1}(i,j,k)}{P(O|\lambda)} = \\ &= \frac{\alpha_t(i,j,k)a_{ijl}b_{ij}(o_{t+1})\beta_{t+1}(i,j,k)}{\sum_{k=1}^{K}\sum_{l=1}^{N^2}\alpha_t(i,j,k)a_{ijl}b_{ij}(o_{t+1})\beta_{t+1}(i,j,k)}\end{aligned} \tag{11}$$

– Define $\gamma(i,j)$ as the probability of being in state i,j at time t, given observation sequence

$$\gamma_t(i,j) = \sum_{l=1}^{N^2}\xi_t(i,j,l). \tag{12}$$

– $\sum_{t=1}^{T}\gamma_t(i,j)$ is the expected number of times state ij is visited
– $\sum_{t=1}^{T-1}\xi_t(i,j,t)$ is the expected number of transition from state ij to l

Update rules:

– $\bar{\pi}_{ijk}$ = expected frequency in state i,j at time $(t=1) = \gamma_1(i,j)$
– $\bar{a}_{ij}$ = (expected number of transition from state i,j to state l)/(expected number of transitions from state i,j):

$$\bar{a}_{ij} = \frac{\sum_t \xi_t(i,j,l)}{\sum_t \gamma_t(i,j)} \tag{13}$$

– $\bar{b}_{ij}(k)$ = (expected number of times in state j and oserving symbol k)/(expected number of times in state j):

$$\bar{b}_{ij}(k)\frac{\sum_{t,o_t=k}\gamma_t(i,j)}{\sum_t \gamma_t(i,j)} \tag{14}$$

4 Experiments

The road signs image database *German Traffic Sign Benchmark* was used in experimenting [1]. As authors wrote [21], the set contains images of more than 1700 traffic sign instances. The size of the traffic signs varies between 1515 and 222193 pixels. The images contain 10% margin (at least 5 pixels) around the traffic sign to allow for the usage of edge detectors. The original size and location of the traffic sign within the image (region of interest, ROI) is preserved in the provided annotations. The images are not necessarily square. Fig. 4 shows the distribution of traffic sign sizes, taking into account the larger of both dimensions of the traffic sign ROI.

Fig. 4. Random representatives of the traffic sign in the GTSRB dataset [1]

Table 1. Comparison of recognition rate

Method	Recognition rate [%]
ESOM [14]	84
HMM [7]	49
1D HMM[our]	81
2D HMM[our]	83

In order to verify the method has been selected fifty objects. Three images for learning and three for testing has been chosen for each object. The 2D HMM has been implemented with parameters $N = 4$, $N^2 = 16$, $K = 16$, $M = 25$. The parameters of HMM was selcted randomly and estimated by modify Baum-Welch algorithm (sec. 4.2). Wavelet transform has been chosen as features extraction technique. Table 1 presents The results of experiments.

5 Conclusion

In this paper, the new conception about road sign recognition with two-dimensional hidden Markov models was presented. We show solutions of principle problems for ergodic 2D HMM, which may be applied for 2D data. Recognition rate of the method is 83%, which is better than 1D HMM. Furthermore, the advantage of this approach is that there is no need to convert the input two-dimensional image on a one-dimensional data and we do not lose the information. The obtained results are satisfactory in comparison to other method and proposed method can be the alternative solution to the others.

References

1. Database German Traffic Sign Benchmark, `http://benchmark.ini.rub.de/Dataset/GTSRB-Final-TrainingImages.zip` (available online April 20, 2014)
2. Eickeler, S., Mller, S., Rigoll, G.: High Performance Face Recognition Using Pseudo 2-D Hidden Markov Models. In: European Control Conference (1999), `http://citeseer.ist.psu.edu`

3. de la Escalera, A., Moreno, L., Salichs, M.A., Armingol, J.M.: Road traffic sign detection and classification. IEEE Transaction Industrial Electronics 44(6), 848–859 (1997)
4. Forney, G.D.: The Viterbi Algorithm. Proc. IEEE 61(3), 268–278 (1973)
5. Garcia-Garrido, M., Sotelo, M., MartinGorostiza, E.: Fast traffic sign detection and recognition under changing lighting conditions. In: Sotelo, M. (ed.) Proceedings of the IEEE ITSC, pp. 811–816 (2006)
6. Gomez-Moreno, H., Lopez-Ferreras, F.: Road-sign detection and recognition based on support vector machines. IEEE Transaction on Intelligent Transportation Systems 8(2), 264–278 (2007)
7. Hsien, J.C., Liou, Y.S., Chen, S.Y.: Road Sign Detection and Recognition Using Hidden Markov Model. Asian Journal of Health and Information Sciences 1(1), 85–100 (2006)
8. Hsu, S.H., Huang, C.L.: Road sign detection and recognition using matching pursuit method. Image and Vision Computing 19, 119–129 (2001)
9. Joshi, D., Li, J., Wang, J.Z.: A computationally Efficient Approach to the estimation of two- and three-dimensional hidden Markov models. IEEE Transactions on Image Processing 15(7), 1871–1886 (2006)
10. Kanungo, T.: Hidden Markov Model Tutorial (1999), `http://www.kanungo.com/software/hmmtut.pdf`
11. Kubanek, M.: Automatic Methods for Determining the Characteristic Points in Face Image. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010, Part I. LNCS (LNAI), vol. 6113, pp. 523–530. Springer, Heidelberg (2010)
12. Li, J., Najmi, A., Gray, R.M.: Image classification by a two dimensional Hidden Markov model. IEEE Transactions on Signal Processing 48, 517–533 (2000)
13. Maldonado-Bascn, S., Acevedo-Rodrguez, J., Lafuente-Arroyo, S., FernndezCaballero, A., Lpez-Ferreras, F.: An optimization on pictogram identification for the road-sign recognition task using SVMs. Computer Vision and Image Understanding 114(3), 373–383 (2010)
14. Nguwi, Y.Y., Cho, S.Y.: Emergent selforganizing feature map for recognizing road sign images. Neural Computing and Application 19, 601–615 (2010)
15. Pazhoumand-dar, H., Yaghoobi, M.: A new approach in road sign recognition based on fast fractal coding. Neural Computing & Application 22, 615–625 (2013)
16. Piccioli, G., Micheli, E.D., Parodi, P., Campani, M.: Robust method for road sign detection and recognition. Image Vision and Computing 14, 209–223 (1996)
17. Prietoa, M.S., Allen, A.R.: Using selforganising maps in the detection and recognition of road signs. Image and Vision Computing 27(6), 673–683 (2009)
18. Rabiner, L.R.: A tutorial on hidden Markov models and selected application in speech recognition. Proc. IEEE 77, 257–285 (1989)
19. Samaria, F., Young, S.: HMM-based Architecture for Face Identification. Image and Vision Computing 12(8), 537–583 (1994)
20. Smorawa, D., Kubanek, M.: Analysis of advanced techniques of image processing based on automatic detection system and road signs recognition. Journal of Applied Mathematics and Computational Mechanics 13(1) (2014)
21. Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C.: The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In: Proceedings of the IEEE International Joint Conference on Neural Networks, pp. 1453–1460 (2011)

22. Vicen-Bueno, R., Gil-Pita, R., Rosa-Zurera, M., Utrilla-Manso, M., López-Ferreras, F.: Multilayer perceptrons applied to traffic sign recognition tasks. In: Cabestany, J., Prieto, A.G., Sandoval, F. (eds.) IWANN 2005. LNCS, vol. 3512, pp. 865–872. Springer, Heidelberg (2005)
23. Vitoantonio Bevilacqua, V., Cariello, L., Carro, G., Daleno, D., Mastronardi, G.: A face recognition system based on Pseudo 2D HMM applied to neural network coefficients. Soft Computing 12(7), 615–621 (2008)
24. Yujian, L.: An analytic solution for estimating two-dimensional hidden Markov models. Applied Mathematics and Computation 185, 810–822 (2007)

Evaluating the Mutual Position of Objects on the Visual Scene Using Morphological Processing and Reasoning

Arkadiusz Cacko and Marcin Iwanowski

Institute of Control and Industrial Electronics, Warsaw University of Technology, ul.Koszykowa 75; 00-662 Warszawa Poland
iwanowski@ee.pw.edu.pl

Abstract. This paper presents methods of extracting spatial relationships between objects on visual scene using directional morphological operations called conic dilation. With additional features describing each object it creates scene description matrix, the structure containing all knowledge of image. Afterward matrix can easily be transformed into Prolog predicates which leads to the inference about scene and possibility of making semi-natural queries about image content.

1 Introduction

In this paper, the image processing methods will be used to extract the information on the position of object within the visual scene. In particular the information on the relative, mutual position of pairs of object will be computed in order to be able to process the information on the composition of objects within scene. The extraction of scene description is divided into two stages. The first step aims at finding the relative position of each object in relation to other objects, using the directional morphological dilation of object with particular, directional structuring element called here, the *conic* dilation. The goal of the second step is to measure the distance between objects. In addition simple individual features of objects are computed in order to get the description of object allowing differentiating one from another. The complete description of visual scene consists thus of relative positions, distances and individual features. All these data will be stored in the scene description matrix that will be further used to extract the verbal description of the scene. The information on the mutual relation of object obtained by the image processing techniques along with the example data referring to the shape and color of object will be stored in a *scene description matrix*. Based on the above martix, the verbal description of visual scene is generated. In order to use the PROLOG reasoning techniques, the verbal scene description may be complemented with some extra informations.

The paper consists of 5 sections. Section 2 presents the previous works on natural language like visual scene description. In section 3, the proposed method for extracting the scene description is given. In section 4, the extraction of PROLOG predicated and reasoning about the scene is described. Section 5 concludes the paper.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_2

2 Previous Works

Knowledge discovery and image understanding are important research field with a great variety of applications. Developing of methods capable for automatically converting images into natural language text sentences using image processing-analysis methods is vital [2].

Paper [14] presents a multi-faceted conceptual framework integrating semantics, texture, and spatial features for automatic image retrieval. It features a high-level representation formalism handling symbolic image descriptions and a unified full-text query framework. In [11,10,8,7] autors are using Prolog language to analize the vision scene. The advantages of using Prolog are its flexibility and simplicity in representation of rules. PrologâĂŹs expressional power arises because it is a Declarative Language and is therefore able to manipulate abstract symbols (words) without explaining what they mean. The ability to apply the reasoning is well suited for the analysis of the scene content. Author notes that combining Image Processing (IP) and Artificial Intelligence (AI) gives many benefits in image knowledge discovery.

3 Extracting the Scene Description

Let $O = O_1 \cup O_2 \cup \ldots O_n$ be the binary image consisting of n connected component (representing image objects), being the result of image segmentation. O_i is thus $i-th$ image object (set of pixels of value '1'). An example of scene consisting of 5 object is shown in Fig. 1a. Starting from this image, the scene description will be extracted in two-stage process that is described below.

3.1 Finding the Relative Position

The relative position of a single object can be estimated only in relation to other objects visible within the vision scene. Assuming that the orientation of the whole scene is fixed, the mutual position will be described using the descriptors based on the division of the complete surrounding of the object into 4 principal sections: top, bottom, left and right. The detection of mutual position is based on morphological dilations [13,5,4] performed with conical structuring elements (conic dilations). The classic definition of dilation is the following:

$$I \oplus B = \bigcup_{b \in B} I_{[-b]}, \tag{1}$$

where $I_{[b]}$ stands for the image (set of pixels) I shifted[1] by vector b. In case of multiple dilations the following relationships holds:

$$(I \oplus B)^{(n)} = (\ldots((I \underbrace{\oplus B) \oplus B) \ldots) \oplus B}_{n \text{ times}} = I \oplus B^{(n)} \text{ where } B^{(n)} = B \oplus \underbrace{B \ldots \oplus B}_{(n-1) \text{ times}} \tag{2}$$

[1] We assume that the origin of the image coordinate system is located in the upper-left image corner.

Let, moreover define the ultimate dilation as:

$$(I \oplus B)^{(\infty)} = (I \oplus B)^{(m)} \text{ where } m = \min_i \left\{ (I \oplus B)^{(i)} = (I \oplus B)^{(i+1)} \right\} \quad (3)$$

Structuring elements applied to dilation in the current study are conical ones. We define the simplest among them in the following manner:

$$\begin{array}{l} \mathcal{C}^B = \{(0,0),(0,-1),(-1,-1),(1,-1)\}, \mathcal{C}^T = \{(0,0),(0,1),(-1,1),(1,1)\}, \\ \mathcal{C}^L = \{(0,0),(-1,0),(-1,-1),(-1,1)\}, \mathcal{C}^R = \{(0,0),(1,0),(1,-1),(1,1)\}. \end{array} \quad (4)$$

Consecutive dilations with the above structuring elements extends the object in one of the principal directions (**B**ottom, **T**op, **L**eft and **R**ight, respectively) creating a 90-degree cone beginning from the object.

Let $\mathcal{C}^D$ be now the structuring element introduced above oriented in one of four principal directions $D \in \{B, T, L, R\}$. Let moreover O be an image under consideration. In order to find the relative position of i-th object O_i comparing with j-th object O_j we need to compute the following intersection:

$$I^D(i,j) = O_i \cap \left(O_j \oplus \mathcal{C}^D\right)^{(\infty)} \quad i \neq j \quad (5)$$

An example of an ultimate dilation is shown in Fig. 1b – ultimately dilated object O_4 (apple) with $\mathcal{C}^B$ structuring element intersects objects O_2,O_3 and O_5. By computing $I^D(i,j)$ for all directions $D \in \{B, T, L, R\}$ we can test whether i-th object is located below (B), above (T), on the right (R) or on the left (L) comparing to j-th one. Namely, if $I^D(i,j) \neq \emptyset$, we can assume that relative position as indicated by D is true. The big advantage is that the result is always correct, even for very complicated arrangement of objects in the scene. Moreover we can compute the confidence of this relation as:

$$c^D_{i,j} = \frac{|I^D(i,j)|}{|O_i|}, \quad (6)$$

where $|.|$ stands for the number of pixels of the argument. In the example from Fig. 1b, the confidence of the relation "is bellow O_4" for O_3 and O_5 equals to 1 ($c^B_{3,4} = c^B_{5,4} = 1$), while $c^B_{5,4} < 1$.

3.2 Distance Computation

In the previous works [10] the distance between objects was computed as a distance between their centers of gravity. This approach is sufficient and give reasonable results for simple shapes or long distance between objects. In the opposite case it does not work properly. An example of such situation is shown in Fig. 2a. The distance between gravity centers of two objects is several times longer than the real distance between objects. In the current study, to get more realistic distance estimation, it is based on the distance from one object to the

(a) (b)

Fig. 1. (a) Examined scene. Objects: 1 - *hat*, 2 - *face*, 3 - *rooster*, 4 - *apple*, 5 - *mouse*, (b) Lower neighbors of *apple* with various conficence level

closest pixel of another object. In order to evaluate such distances, the distance functions are computed as:

$$d(O_i)[p] = \begin{cases} \min_{q \in O_i} \{dist(p,q)\} & \text{if } p \notin O_i, \\ 0 & \text{if } p \in O_i, \end{cases} \tag{7}$$

where $dist(p,q)$ stands for the distance between pixels p and q. The above distance function can the thresholded at given level k to obtain the surrounding of O_i of the radius k.

Based on the above distance function two measures[2] describing the mutual relation between objects O_i and O_j can be computed:

$$\begin{aligned} d_{i,j} &= \min_k \{(d(O_i) < k) \cap O_j \neq \emptyset\}, \\ d'_{i,j} &= \min_k \{(d(O_i) < k) \cap O_j = O_j\}, \end{aligned} \tag{8}$$

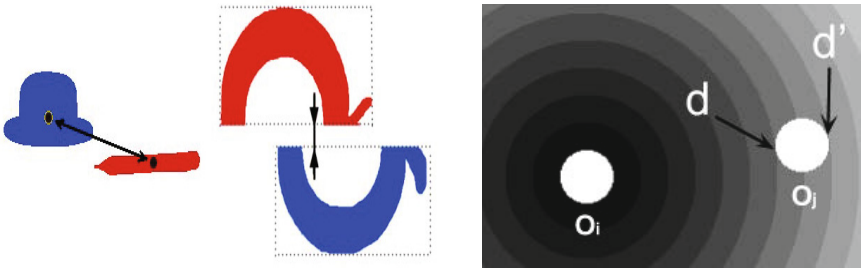
where $d(O_i) < k$ stands for the surrounding of O_i or the radius k. The measure $d_{i,j}$ represents the distance between object O_i and its closest pixel belonging to O_j, while the measure $d'_{i,j}$ stand for the distance between object O_i and its farthest pixels belonging to O_j – for an example see Fig. 2b.

3.3 Individual Features

In order to find the properties of individual object, in the current study the following measures are computed for all objects O_i:

1. perimeter p – lenght of the boundary of the region,
2. euler number e – number of objects in the region minus the number of holes in those objects,

[2] This measures formally are not distances because they are, in general case, not symmetric.



(a) (b)

Fig. 2. (a) Gravity and boundig box distance, (b) Distance function d and d'

3. area a – the actual number of pixels in the region,
4. hue h – color of object.

The above features are just exemplary ones. In fact all features used in pattern recognition may be used, depending on the complexity and properties of object being considered.

3.4 Scene Description Matrix

All above descriptors are stored is a data structure which will be called scene description matricx. For n objects, the description matrix R is $n \times n$ matrix, where $R(i,j)$ for $i \neq j$ describes relations between i-th and j-th object obtained from conic dilations and distance functions:

$$[c^B_{i,j}; c^T_{i,j}; c^L_{i,j}; c^R_{i,j}; d_{i,j}; d'_{i,j}] \tag{9}$$

Elements $R(i,j)$ contain individual features of object as vector [**h**ue, **e**uler number, **a**rea, **p**rimeter] - $[h_i, e_i, a_i, p_i]$.

An example matrix of the image shown in Fig. 1a is given in Table 1.

Table 1. Scene description matrix

	1	2	3	4	5
1	**[yellow,1, 5814,341]**	[1;0;0.5; 0.3;12;165]	[0.07;0; 0;1;85;165]	[0;0;0; 1;246;350]	[0;0;0; 1;280;320]
2	[0;1;0.3; 0.2;12;77]	**[green,1 10208, 416]**	[0;0.05;0; 1;93;170]	[0;0.3;0; 1;292;396]	[0;0;0; 1;324;366]
3	[0;0.04;1; 0;85;204]	[0.02;0;1; 0;93;185]	**[violet,1 3334,371]**	[0;0;0; 0.7;146;260]	[0;1;0; 0.1;175;221]
4	[0;0;1; 0;246;366]	[0.17;0; 1;0;292;403]	[1;0;1; 0;146;224]	**[red,0 5646,423]**	[1;1;1; 1;6;30]
5	[0;0;1; 0;280;402]	[0;0;1; 0;324;434]	[0.9;0;0.7; 0;174;253]	[0.4;0.4;0.4; 0.4;6;44]	**[grey,0 868,193]**

4 Reasoning-Based Processing

4.1 Obtaining the Predicate-Based Description

Based on the relationship matrix obtained in the previous step the predicates describing the properties of scene objects are extracted. All information in matrix can be translated to Prolog predicates using the below algorithm:

Algorithm 1. Predicates extraction algorithm

Data: NxN description Matrix
for $object \leftarrow 1$ **to** N; ***for*** $neighbor \leftarrow 1$ **to** N **do**
  if $object \neq neighbor$ **then**
    for $D \leftarrow B, T, L, R$ **do**
      if $c^{D}_{object,neighbor} > threshold$ **then**
        $direction$ = 'bottom', 'top', 'left' or 'right' depending on D
        addFact(pos($object$,$neighbor$,$direction$,$c_{object,neighbor}$))
        addFact(distD($object$,$neighbor$,$d_{object,neighbor}$))
        addFact(distDp($object$,$neighbor$,$d'_{object,neighbor}$))
  else
    addFact(feat($object$, h_{object}, e_{object}, a_{object}, p_{object}))

Example output for object 2:
pos(2,1,top,1) - object 1 is over 2 with confidence 1
pos(2,1,left,0.3), pos(2,1,right,0.2), pos(2,3,top,0.05), pos(2,3,right,1),
pos(2,4,top,0.3), pos(2,4,right,1), pos(2,5,right,1)
distD(2,1,12) - distance d from 1 to 2 equals 12 pixels
distD(2,3,93), distD(2,4,292), distD(2,5,324)
distDp(2,1,77) ...
feat(2,green,1,10208,416) - object 2 with values of individual features

According to this algorithm, the relative position in particular direction is defined only if the confidence is greater than given *threshold*. Thanks to this assumption only the objects with most of their area included in the result of the conic dilation of another object are considered.

Data extracted in this way contain the basic knowledge of the scene – contain the previously described spatial relationships ('pos','distD' and 'distDp' predicates) between all objects and the individual features ('feat' predicates) of each object.

4.2 Reasoning about Image Content

The predicates describing the given visual scene are the base text descriptors of the scene. Starting from them, more complex descriptors are formulated in order to describe more sophisticated inter-object relations. Using Prolog inference system - many other relationships can be discovered from the basic relationship descriptors extracted directly from description matrix. Using previously discovered simple spatial relationships can easily find objects that are e.g. between other objects, are in their center or lie diagonally. There is a lot more complex

dependencies, which can be obtained by composition of fundamental ones. Language Prolog helps in discovering them through the ability to create rules of inference e.g. such as:

- middle(X,Z,Y):-pos(X,Z,down,_),pos(Y,Z,up,_).
- contains(X,Y):-pos(X,Y,down,1),pos(X,Y,up,1),pos(X,Y,left,1),pos(X,Y,right,1).
- isClose(X,Y):-distD(X,Y,Z<50).
- diagonally(X,Y):-pos(X,Y,down,_),pos(X,Y,left,_)
- ...

In this way many complex relationships can be found which may than be used to enhance description of visual scene.

With such structured data we have ability to make a query about this visual scene like:

Is there any green object close to another?	feat(X,green,_,_),isClose(X,_).
How many object contains another object?	findall(N,contains(X,_), Ns), length(Ns, X).
Which objects are diagonaly to blue object?	diagonally(X,Y),feat(X,blue,_,_).
Which big objects are in the middle of 1 and 5 object?	middle(1,Y,5),feat(Y,_,K>100,_).
Which object is the biggest on the scene?	pos(Max,Y,_,_),\+((pos(X,_,_,_), $X < Max$)).
How far from the 1 obj. are other obj.?	distD(1,X,Dist).

Using Prolog we could create any question about facts which were extracted from image. Thanks to it there is possibility to produce answears to semi-natural language questions about scene.

5 Conclusions

Methods for spatial describe of visual scene have been proposed. Scene description matrix containing basic dependaces between objects is builded using morphological operations. Conic dilation produce always correct directional relations between pairs of objects even in very complicated scene. This informations supplemented by object features creates scene description matrix.

After matrix creation data can be trensformed into Prolog predicates. It gives possibility to generate complex verbal scene description based on reasoning. This also allows creating a semi-natural language questions about content of visual scene.

References

1. Kuo, Y.H.: Unsupervised semantic feature discovery for image object retrieval and tag refinement. IEEE Transactions on Multimedia 14(4), 1079–1090 (2012)
2. Bourbakis, N.: Image understanding for converting images into natural language text sentences. In: 2010 International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE), vol. 1 (2010)

3. Lyu, M.R.T., Ma, H., Zhu, J., King, I.: Bridging the semantic gap between image contents and tags. IEEE Trans. Multimedia 12(2), 462–473 (2010)
4. Iwanowski, M.: Metody morfologiczne w przetwarzaniu obrazów cyfrowych, EXIT (2009)
5. Soille, P.: Morphological image analysis. Springer (1999, 2004)
6. Mojsilovic, B., Rogowitz, A.: Capturing image semantics with low-level descriptors. In: Proceedings of the 2001 International Conference on Image Processing, vol. 1, pp. 18–21 (2001)
7. Batchelor, B.G., Whelan, P.F.: Intelligent Vision Systems for Industry. Springer, London (1997)
8. Batchelor, B.G., Jones, A.C.: PIP - An Integrated Prolog Image Processing Environment. In: Prolog for Industry: Proceedings of the LPA Prolog Day at the RSA. Logic Programming Associates Ltd., London (1995)
9. Jones, A.C.: Image Processing in Prolog: Getting the Paradigm Right. The ALP Newsletter 4, 8 (1995)
10. Batchelor, B.G., Waltz, F.: Interactive Image Processing for Machine Vision. Springer, London (1993)
11. Batchelor, B.G.: Intelligent Image Processing in Prolog. Springer, London (1991)
12. Biederman, I.: Recognition-by-components: A theory of human image understanding. Psychol. Rev. 94(2), 115–147 (1987)
13. Serra, J.: Image analysis and mathematical morphology, vol. 1. Academic Press (1983)
14. Belkhatir, M.: Unifying multiple description facets for symbolic image retrieval. In: IEEE International Conference on Image Processing, ICIP 2005, vol. 3, pp. III:189–III:192 (2005)

Vascular Biometry

Ryszard S. Choraś

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

Abstract. Biometric measures have been used to identify people based on feature vectors derived from their physiological/behavioral characteristics. This paper proposes a method of human biometric identification based on human body organs vasculature. The blood vessels of several part of body have a unique pattern. Feature vectors based on linear patterns such as blood vessel are suitable to identification of an individual. This paper focused on detecting and analyzing vasculature of finger. The proposed algorithm using steerable filters and global and local features.

1 Introduction

All biometric systems work in a similar fashion:

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

Biometric systems have four main components [5]: data acquisition (sensor's), feature extraction, biometric database, matching-score and decision-making modules (Fig. 1). The input subsystem consists of a special sensor needed to acquire the biometric signal. Invariant features are extracted from the signal for representation purposes in the feature extraction subsystem. During the enrollment process, a representation (called template) of the biometrics in terms of these features is stored in the system. The matching subsystem accepts query and reference templates and returns the degree of match or mismatch as a score, i.e., a similarity measure. A final decision step compares the score to a decision threshold to deem the comparison a match or non-match.

In generally, a biometric system is a pattern recognition system that recognizes a person on the basis of a feature vector derived from a specific physiological or behavioral characteristic that the person possesses. Physiological Biometrics - also known as static biometrics - based on data derived from the measurement of a part of a person's anatomy. For example, fingerprints and iris patterns, as well as facial features, hand geometry and retinal blood vessels. Behavioral

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_3

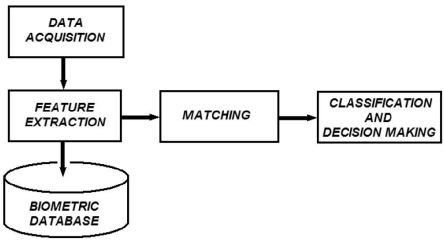


Fig. 1. Biometric system

biometrics based on data derived from measurement of an action performed by a person and, distinctively, incorporating time as a metric, that is, the measured action. The behavioral characteristics measure the movement of a user, when users walk, speak, type on a keyboard or sign their name.

The problem of vein extraction arises mainly in biometric and biomedical applications. The advantages od vein (vascular) biometry lie in the following aspects:

1. the vein image acquisition maybe is non contact,
2. vein recognition belongs to the kind of live body identification,
3. vein pattern is an internal feature and difficult to forge.

Biometric vein system is based on the fact that the vein pattern is distinctive for various individuals. The human veins are located inside the human body. Therefore, the recognition system will capture images of the vein patterns inside of humans parts of body (e.g. fingers) by applying light transmission to this human organs (Fig. 2).

The vessels are piecewise linear and can be represented by many connected line segments with direction and grey level continuous. Results from scanning these body parts are specific blood vessel linear patterns with significant points across images. Feature vector extracted around significant points is uses to distinguish blood vessel.

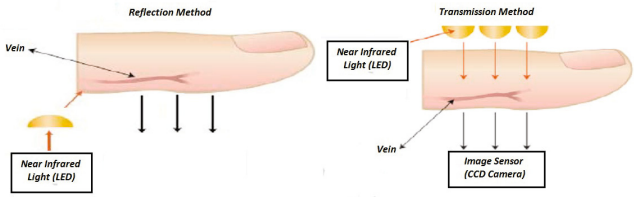
Typically vessel recognition starting with detection of junctions points. The locally vessel structure around these points contain information which is used to construct the feature vector.

The paper consists of four sections. Section 2 explains the complete proposed system. Section 3 contains experimental results. Section 4 provides conclusion.

2 Processing - Extraction of Vein Structure

There are few main parts for the vein recognition process:

– Vein image acquisition,
– Vein image processing,



(a) (b)

Fig. 2. Transmission method: (a) Light reflection, (b) top light transmission

– Feature extraction.
– Matching.

The diagram of the processing stages implemented is shown in Fig. 3 and each stage is discussed.

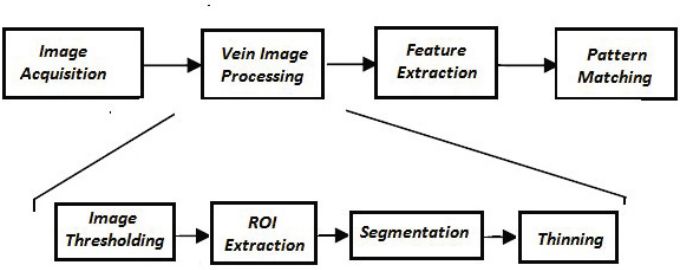
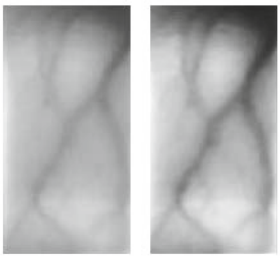


Fig. 3. Processing diagram

Vein images obtained cannot be used for identification directly. Firstly, we perform grey scale normalization, as follows:

$$f_{out}(x,y) = 255 \times \left(\frac{f_{in}(x,y) - min}{max - min} \right)^{\gamma} \tag{1}$$

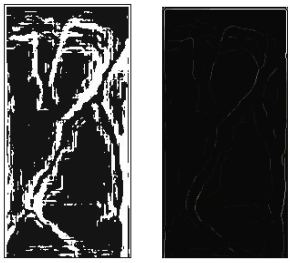
$f_{out}(x,y)$ is the color level for the output pixel (x,y) after the contrast stretching process. $f_{in}(x,y)$ is the color level input for data the pixel (x,y). max - is the maximum value for color level in the input image. min - is the minimum value for color level in the input image, γ - constant that defines the shape of the stretching curve.



(a) (b)

Fig. 4. Finger vein images: (a) oryginal image, (b) image after gray scale normalization

To obtain the vein binary image several alternatives method can be use from morphological to multi-resolution analysis methods. We use the typical adaptive threshold algorithm.



(a) (b)

Fig. 5. Finger vein images: (a) oryginal image after threshold operation, (b) image after thinning operation

3 Steerable Filters

The steerable filters are used for detection of characteristic vein lines.
Steerable filters are a class oriented filters in which any filter is represented by a linear combination of set of basis filters.

Let a 2D Gaussian function $g(x, y)$ is defined as

$$g(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} \tag{2}$$

The first derivative with respect to x is defined as the following equation:

$$g_x^{'}(x,y) = \frac{\partial}{\partial x}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} = -\frac{1}{\sqrt{2\pi}\sigma^3}\, x\, e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} \tag{3}$$

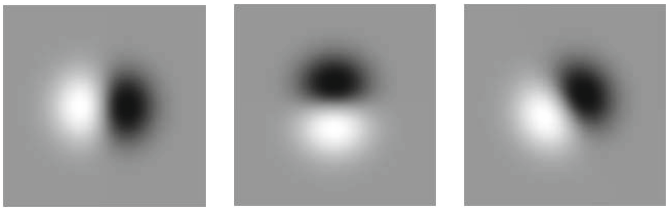
Above equation rotated by angle $\theta = 90$, it is the same as the derivation of with respect to y

$$g_y^{'}(x,y) = g_{x(\theta=90)}^{'}(x,y) = \frac{\partial}{\partial y}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} = -\frac{1}{\sqrt{2\pi}\sigma^3}\ y\ e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} \quad (4)$$

The first derivative of $g(x,y)$ rotated by any angle $\theta = 90$ can be obtained by

$$g_\theta^{'}(x,y) = \frac{\partial}{\partial\theta}g(r,\theta) = -\frac{1}{\sqrt{2\pi}\sigma^3}(x\cos\theta + y\sin\theta)e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} = \\ = \cos\theta g_x^{'}(x,y) + \sin\theta g_{x(\theta=90)}^{'}(x,y) \quad (5)$$

Since $g_x^{'}(x,y)$ and $g_y^{'}(x,y)$ span the set of $g_\theta^{'}(x,y)$ filters we call them basis filters for $g_\theta^{'}(x,y)$. $\cos\theta$ and $\sin\theta$ terms are the corresponding functions for those basis filters [1].



(a) (b) (c)

Fig. 6. A first order Gaussian derivatives: (a) $\frac{\partial g}{\partial x}$, (b) $\frac{\partial g}{\partial y}$, (c) $\cos(30^\circ)\frac{\partial g}{\partial x} + \sin(30^\circ)\frac{\partial g}{\partial y}$

Based on second derivatives od 2D Gaussian function we have the following steerable filters

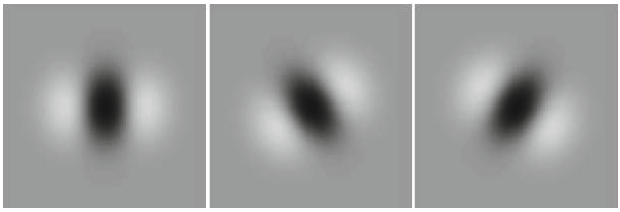
$$g_\theta^{''}(x,y) = \cos^2\theta g_{xx}^{''}(x,y) + \sin^2\theta g_{yy}^{''}(x,y) - 2g_{xy}^{''}(x,y)\cos\theta\sin\theta \quad (6)$$

For a given σ, convolution of an image with three filter kernels appears to be sufficient to determining vein orientation. Equation (7) in the orientation $\theta = \frac{n\pi}{3}; n = 0,1,3$ can be expressed as

$$g_\theta^{''}(x,y) = \frac{1}{3}(1+2\cos(2\theta))g_{0^\circ}^{''}(x,y) + \frac{1}{3}(1+2\cos(2\theta)+\sqrt{3}\sin(2\theta))g_{60^\circ}^{''}(x,y) \\ +\frac{1}{3}(1+2\cos(2\theta)-\sqrt{3}\sin(2\theta))g_{120^\circ}^{''}(x,y) \quad (7)$$

The dominant orientation at each pixels is given by

$$\theta(x,y) = \arctan\left[\frac{\sqrt{3}(g_{60^\circ}^{''}(x,y) - g_{120^\circ}^{''}(x,y))}{2g_{0^\circ}^{''}(x,y) - g_{60^\circ}^{''}(x,y) - g_{120^\circ}^{''}(x,y)}\right] \quad (8)$$



(a) (b) (c)

Fig. 7. Three directional second order Gaussian derivatives: (a) $\theta = 0°$, (b) $\theta = 60°$, (c) $\theta = 120°$

The information captured by the filter depends on the width of the Gaussian function. For larger σ (wider Gaussian function), filter responds to lower frequency. Based on eq. (7) the orientation with max output can be determined. The responds $g''_\theta(x, y)$ at the selected scale is stored for each pixel and large values are allow to further processing.

Fig. 8. Filter output (second order Gaussian derivatives) for finger vein image

4 Radon Transform

The Radon transform of a 2-D function $g''_\theta(x, y)$, denoted as $R(r, \theta)$, is defined as its line integral along a line inclined at an angle θ from the y-axis and at a distance r from the origin (as shown in Fig. 9). In other words, $R(r, \theta)$ is the 1-D projection of $g''_\theta(x, y)$ at an angle θ.

From Radon transform, only the coefficients greater than the appropriate threshold:

$$th = \mu + \frac{(max - \mu)}{\alpha} \tag{9}$$

where α is the adaptive factor, μ is the mean of Radon coefficients and max is the maximum of the Radon coefficients along each projection.

Points in (RT) threshold image determine central point of ROI in spatial space. The size of each ROI block in our application is $k \times l \quad (k = l = 20)$.

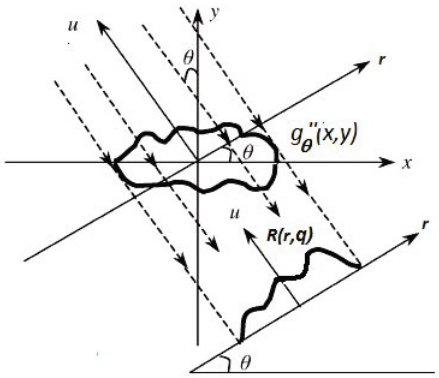


Fig. 9. The Radon transform $R(r,\theta)$ is the 1-D projection of $g''_\theta(x,y)$ at an angle θ

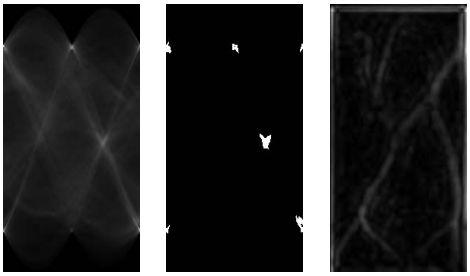


Fig. 10. Radon transform (RT) of images $g''_\theta(x,y)$, threshold version of (RT) and reconstructed image

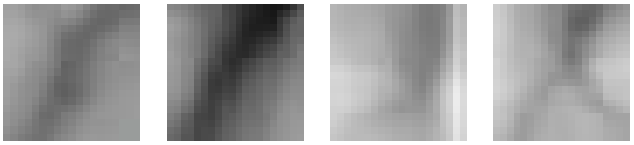


Fig. 11. ROI images determine by threshold version of (RT)

For each block (Fig. 11) are computed some texture parameters

$$energy(x,y) = \frac{1}{20 \times 20} \sum_{k=1}^{20} \sum_{l=1}^{20} (g_\theta^{''(ROIi)}(x,y))^2 \tag{10}$$

$$entropy(x,y) = \frac{1}{20 \times 20} \sum_{k=1}^{20} \sum_{l=1}^{20} g_\theta^{''(ROIi)}(x,y)(-\ln g_\theta^{''(ROIi)}(x,y)) \tag{11}$$

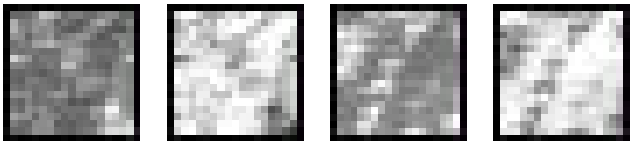


Fig. 12. ROI Energy and Entropy images: (a) ROI1; (b) ROI2

The feature vector is constructed using $energy(x, y)$ and $entropy(x, y)$ as feature components.

5 Conclusion

A new method for recognition vein images has been presented. This method based on steerable, Radon and statistical finger vein features. This paper analysis the details of the proposed method. These features can be used to personal identification. Experimental results have demonstrated that this approach is promising to improve person identification.

References

1. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. IEEE Trans. on Pattern Analysis and Machine Intelligence 13(9), 891–906 (1991)
2. Jacob, M., Unser, M.: Design of steerable filters for feature detection using Canny-like criteria. IEEE Trans. on Pattern Analysis and Machine Intelligence 26(8), 1007–1019 (2004)
3. Kirbas, C., Quek, K.: Vessel extraction techniques and algorithm: a survey. In: Proceedings of the 3rd IEEE Symposium on BioInfomratics and Bioengineering (BIBE 2003) (2003)
4. Jain, A.K., Bolle, R., Pankanti, S.: Biometrics.: Personal Identification in Networked Society. Kulwer Academic (1999)

Clustering-Based Retrieval of Similar Outfits Based on Clothes Visual Characteristics

Piotr Czapiewski[1], Paweł Forczmański[1],
Dariusz Frejlichowski[1], and Radosław Hofman[2]

[1] West Pomeranian University of Technology, Szczecin,
Faculty of Computer Science and Information Technology,
Żołnierska 49, 71–210 Szczecin, Poland
{pczapiewski,pforczmanski,dfrejlichowski}@wi.zut.edu.pl
[2] FireFrog Media sp. z o.o., Poznań, Poland

Abstract. The fashion domain has been one of the most growing areas of e-commerce, hence the issue of facilitating cloth searching in fashion-related websites becomes an important topic of research. The paper deals with searching for similar outfits in the clothing images database, using information extracted from unconstrained images containing human silhouettes. Medoids-based clustering is introduced in order to detect groups of similar outfits and speed up the retrieval procedure. Exemplary results of experiments performed on real clothing datasets are presented.

1 Introduction

1.1 Motivation

Recently the fashion domain has been one of the most dynamically growing areas of e-commerce and social networking. Given the popularity of the topic, especially among women, number of potential customers is vast. At the same time, buying clothes on-line poses a bigger problem, than in other areas of e-commerce. First, lack of physical contact with the merchandise is for many people a discouraging factor. Second, due to the great variety of clothing styles and certain difficulties in clearly and unambiguously describing them, it is quite difficult to search for desirable clothes. Furthermore, in real-life shopping for clothes it is quite common to ask for advice of a shopping assistant.

Due to the above factors, the analysis and representation of clothes became an important topic of research within several subareas of computer science, including computer vision, knowledge representation, information retrieval and recommender systems. The approach presented in this paper is intended to be applied within a fashion recommender system. The ultimate goal of such a system would be to learn the individual style of the user and then to recommend some matching new clothes. An important part of such a system is comparing clothes in terms of their visual characteristics and retrieving outfits similar to a given pattern from the database.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_4

The proposed approach consists in analysing the images of human silhouettes, extracting segments corresponding to particular body parts (and as such, to particular garments) and calculating a certain colour-based visual descriptors for them. Then, the clustering of outfits is performed, in order to facilitate similar outfits retrieval and — potentially — to discover recurring fashion styles. The similarity measure utilized for clustering takes into consideration the whole outfit, not just a single garment at the time.

The rest of the paper is organized as follows. Section 1.2 presents the relevant previous research. Section 2 presents the details of the proposed approach, describing subsequently: feature extraction stage, the proposed similarity measure, clustering and retrieval procedures. Section 3 contains results of experiments performed on real clothing datasets.

1.2 Previous Work

Analysing various fashion-related data just recently became an active area of research, especially within the field of computer vision, focusing on analysing images containing clothes and outfits.

Zhang et al. [14] proposed a human-computer interaction system called a responsive mirror, intended to be used as an interactive tool supporting shopping in a real retail store. Cloth type and some attributes are automatically extracted from the image and similar clothes are looked for in a database. Di et al. [4] proposed a method to recognize clothes attributes from images, limiting the area of interest to one category of clothes (coats and jackets). A similarly limited approach can be found in [2], where only upper body clothes are analysed. None of the above deals with the whole outfits, only with single items of clothing.

The topic of similarity between items of clothing from computer vision perspective has been dealt with in [3]. In [9] authors deal with cloth similarity in terms of the semantic description, without considering any visual descriptors. Both above mentioned papers deal only with single elements of clothing, no means of analysing complete outfits is proposed.

To our best knowledge, no research was reported aimed at clustering clothing imagery data containing complete outfits.

2 Method Description

2.1 Feature Extraction

The algorithm used to extract outfit's visual characteristics from the image containing human silhouette has been described in details in [7]. It is based on detecting faces using Viola-Jones approach [12] and on methods for detecting human pose proposed in [15,11,10,5,6]. As a result the segments of the image are indicated containing particular body parts (Fig. 1).

After obtaining the segments corresponding to particular body parts and, hence, to particular garments, the visual descriptors are calculated for each segment. A simplified dominant colour descriptor was used, based on the one defined within MPEG-7 standard [13].

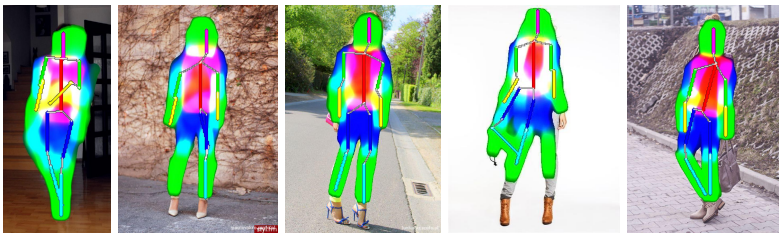


Fig. 1. Example of human pose detection and body parts segmentation

For each of 10 predefined body segments (face, torso, upper arms, lower arms, thighs and lower legs) the descriptor is calculated according to the following procedure [7]:

- after conversion to HSV and colour quantization, a 72-bin histogram is created;
- 8 bins with the highest values are preserved, the others are zeroed;
- the histogram is normalized.

The above procedure leads to obtaining a 720 element feature vector, which is further used for similarity estimation and clustering.

2.2 Similarity Measure

In order to compare visual characteristics of outfits the following similarity measure was proposed in [7]. Given that not all elements of the outfit are equally significant when assessing similarity, the weights have been assigned to particular segments:

$$w = (0.40.10.10.10.10.050.050.050.050), \tag{1}$$

where the subsequent weights correspond to torso, upper arms, thighs, lower arms, lower legs and face. Given the feature vector described in the previous section and the above weights vector, the similarity measure between two human silhouettes H_A and H_B is calculated as:

$$D(A,B) = \sum_{k=1}^{10} \sum_{i=1}^{72} w_k \cdot min(H_A(i,k), H_B(i,k)). \tag{2}$$

where $H_x(i,k)$ denotes the k-th bin of the histogram corresponding to the i-th body segment. The value of D lies within $< 0,1 >$ interval, the value 1 being equivalent to high similarity.

2.3 Clustering

The key element of the approach proposed here is performing cluster analysis based on the above defined features. The underlying assumption is that in any

large enough clothing database there should naturally occur certain groups of outfits, that are similar to each other in terms of colours composition. This could be used for two major purposes: speeding up similar outfits retrieval and analysing individual styles within the fashion recommender system.

Given the similarity measure $D(A, B)$, the following distance measure can be naturally introduced:

$$d(A, B) = 1 - D(A, B). \tag{3}$$

Given the form of the introduced feature vector and weighted similarity measure, no centroid-based methods (e.g. k-means algorithm) can be used for clustering. Hence, the k-medoids algorithm [8] was used, as it is able to partition data based solely on the distance measure. During our experiments the implementation developed by Brookings et al. [1] was used.

As a result of clustering, each outfit from the database is assigned to one of k clusters, represented by single outfits from the database (medoids). The number of clusters was determined experimentally, by analysing two factors: the silhouettes of obtained clusters and the retrieval accuracy obtained for different k values.

2.4 Similar Outfits Retrieval

The initial concept of similar outfits retrieval, as presented in [7], assumed calculating similarity of the given reference outfit to all the outfits in the database. Such an approach guarantees finding the best match (in terms of selected similarity measure), but it is far from being optimal in terms of computation requirements. Hence we propose to use the cluster information as means of speeding up the retrieval.

The procedure of retrieving n most similar outfits was implemented as follows:

- given the reference outfit A, calculate the feature vector H_A;
- for each medoid M_i, $i = 1...k$, calculate the distance measure $d(A, M_i)$;
- find the medoid M_c which is closest to A;
- for each outfit O_j in cluster c (represented by medoid M_c) calculate the distance $d(A, O_j)$;
- return n closest outfits from the cluster c.

As a result, the average number of comparisons between outfits required during retrieval is expected to drop by a factor of k.

3 Experimental Results

For the experiments a dataset of 500 women outfits was used. The images were gathered from various websites and social networks. Out of these, for 417 images it was possible to extract human body segments, and this set was used in clustering experiment.

Fig. 2. Example of obtained clusters of outfits

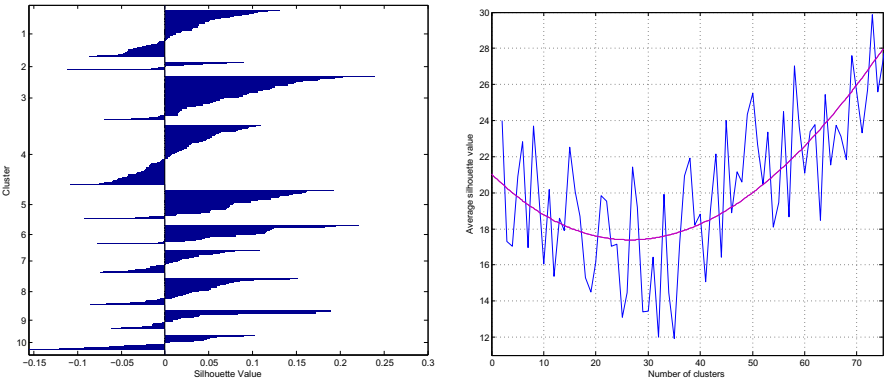


Fig. 3. Analysis of cluster silhouettes: all silhouettes for $k = 10$ (left) and average silhouette value depending on k (right)

The initial analysis showed, that certain groups of similar outfits can be found. An example of similar outfits forming a cluster is shown in Fig. 2). However, the clusters are not very distinct. As can be seen from exemplary silhouette plot (Fig. 3), not all the outfits fit unambiguously to one definite cluster.

In order to find the optimal number of clusters for a test dataset, the clustering was performed repeatedly with varying values of k. Then, the silhouette was calculated for each data point. The average value of the silhouettes was taken as a measure of clustering validity. The plot of average silhouette values for different values of k is presented in Fig. 3.

As can be seen in Fig. 3, the minimum average silhouette value occurs for $k = 25$. However, a high variability can be observed.

At this moment it is crucial to consider the motivation for clustering these data. The ultimate goal is retrieving similar outfits — let us consider two retrieval scenarios: A) looking for similar items in the whole database, B) searching only within the closest cluster. Should the perfectly distinct clusters appear in data, then the retrieval results would be exactly the same in those two scenarios. However, if clustering is not as clear and distinct, then for some items the result of retrieval would be different. It can be expected, that the larger value of k, the larger the difference between both scenarios' results (since less items fall into any given cluster).

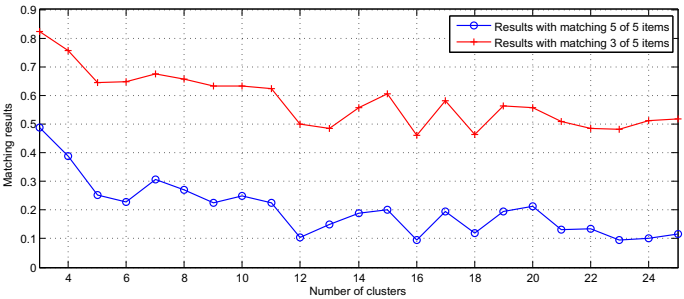


Fig. 4. Comparison between cluster-based and full database retrieval results depending on number of clusters

Fig. 5. Examples of cluster-based similar outfit retrieval: reference outfit (first column) and retrieval results

Given the above, the following experiment was conducted. For each item in test dataset the 5 most similar outfits were retrieved — first according to scenario A (whole database), then to scenario B (within the closest cluster).

The 5 items sets retrieved in both cases where compared and the coefficient was calculated indicating, how close those two results match:

$$\delta(i) = \frac{\#(R_A \cap R_B)}{n}, \tag{4}$$

where R_A and R_B denote resulting sets from scenarios A and B respectively, # denotes set cardinality and n denotes the number of similar outfits sought (here: 5). Next, the histogram of $\delta(i)$ values was created, and the number matches was analysed. In Fig. 4 a plot is presented illustrating, how many perfect matches occurred for different values of k (lower line) and how many results with at least 3 matching items (upper line).

As can be seen from Fig. 1, a significant decrease in retrieval accuracy occurs for number of clusters greater than 10. For $k = 10$ more than 60

Finally, the value $k = 10$ was selected as a number of clusters for the examined dataset. Exemplary results of retrieving similar outfits are presented in Fig. 5.

4 Conclusions

The method presented in the paper allows for retrieval of similar outfits from the database of clothing images. The visual characteristics of the whole outfit is taken into consideration, not only of a single garment. The retrieval can be accelerated by introducing initial clusterization of outfits in the database. Presented experimental results prove the effectiveness of searching for similar cloths using cluster-based approach.

Further research is ongoing in order to incorporate more factors into similarity estimation, including texture information and semantic description. Also, the effectiveness of cluster-based retrieval can be further enhanced by introducing some form of fuzzy clustering.

Acknowledgements. The project "Construction of innovative recommendation based on users' styles system prototype: FireStyle" (original title: "Zbudowanie prototypu innowacyjnego systemu rekomendacji zgodnych ze stylami uzytkowników: FireStyle") is the project co-founded by European Union (project number: UDA-POIG.01.04.00-30-196/12, value: 14.949.474,00 PLN, EU contribution: 7.879.581,50 PLN, realization period: 01.2013-10.2014). European funds — for the development of innovative economy (Fundusze Europejskie — dla rozwoju innowacyjnej gospodarki).

References

1. Brookings, T., Grashow, R., Marder, E.: Statistics of neuronal identification with open and closed loop measures of intrinsic excitability. Frontiers in Neural Circuits 6(19) (2012)

2. Chen, H., Gallagher, A., Girod, B.: Describing clothing by semantic attributes. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 609–623. Springer, Heidelberg (2012)
3. Chen, Q., Li, J., Liu, Z., Lu, G., Bi, X., Wang, B.: Measuring clothing image similarity with bundled features. International Journal of Clothing Science and Technology 25(2), 119–130 (2013)
4. Di, W., Wah, C., Bhardwaj, A., Piramuthu, R., Sundaresan, N.: Style finder: Fine-grained clothing style detection and retrieval. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 8–13. IEEE (2013)
5. Eichner, M., Ferrari, V.: Better appearance models for pictorial structures (2009)
6. Eichner, M., Marin-Jimenez, M., Zisserman, A., Ferrari, V.: 2d articulated human pose estimation and retrieval in (almost) unconstrained still images. International Journal of Computer Vision 99(2), 190–214 (2012)
7. Forczmański, P., Frejlichowski, D., Czapiewski, P., Okarma, K., Hofman, R.: Comparing clothing styles by means of computer vision methods. In: Proceedings of International Conference on Computer Vision and Graphics, ICCVG 2014 (2014)
8. Kaufman, L., Rousseeuw, P.: Clustering by means of medoids. North-Holland (1987)
9. Liu, Z., Wang, J., Chen, Q., Lu, G.: Clothing similarity computation based on tlac. International Journal of Clothing Science and Technology 24(4), 273–286 (2012)
10. Ramanan, D.: Part-based models for finding people and estimating their pose. In: Visual Analysis of Humans, pp. 199–223. Springer (2011)
11. Ramanan, D., Forsyth, D.A., Zisserman, A.: Tracking people by learning their appearance. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(1), 65–81 (2007)
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. I–511. IEEE (2001)
13. Yamada, A., Pickering, M., Jeannin, S., Jens, L.: Mpeg-7 visual part of experimentation model version 9.0-part 3 dominant color, iso. Technical report, IEC JTC1/SC29/WG11 (2001)
14. Zhang, W., Begole, B., Chu, M., Liu, J., Yee, N.: Real-time clothes comparison based on multi-view vision. In: Second ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC 2008, pp. 1–10. IEEE (2008)
15. Zisserman, A.: Human pose estimation in images and videos (2010)

Improving Shape Retrieval and Classification Rates through Low-Dimensional Features Fusion

Paweł Forczmański

West Pomeranian University of Technology, Szczecin,
Faculty of Computer Science and Information Technology,
Żołnierska Str. 52, 71–210 Szczecin, Poland
pforczmanski@wi.zut.edu.pl

Abstract. In the paper an approach to shape classification and shape retrieval is described. Although, most of available shape descriptors give a very good recognition accuracy or retrieval rate, they suffer from one serious limitation, namely, they do not take into account the dimensionality of feature space, hence the computational costs of similarity evaluation is rather high. The problem occurs often in the hardware implementations, where the complexity of processed data should be minimized. Hence we propose a method of joining low-dimensional feature vectors derived from shapes to increase the retrieval rate and classification accuracy.

1 Introduction

Visual information plays an important role in human perception. It may be represented by various forms, such as still images, video streams, simplified graphics, 3D models, animation etc. One of the basic visual information forms to be processed is image, since the need of retrieving a desired image from some collection is shared by ordinary users as well as many professional groups, e.g. journalists, commercial designers, art historians. While it is very attractive to provide higher level query using indexing methods such as keyword (tag) indexing and textual annotation to make use of facilitation of query language, such as SQL, from current database techniques, there are several drawbacks with these indexing methods [1]: (i) they do not conform to a standard description language, (ii) they are inconsistent, (iii) they are subjective, i.e. they might not capture the image content and (iv) they are time consuming. In order to overcome these drawbacks, most directions in research on image retrieval focuses on Content-Based Image Retrieval (CBIR), which utilizes low-level image features such as color, texture and shape. A few dozens of commercial and academic prototypes of CBIR systems have been developed for last 15 years to allow searching through image databases by image content. However, the problem of CBIR remains still unsolved.

This article is organized as follows: First, we introduce shape as object descriptor, then we present selected descriptors and discuss a method of joining them into single feature vector. Further in the paper we present the results of experiments performed on MPEG-7 benchmark database and in the end we summarize our work.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_5

1.1 Shape as a Descriptor

Shape (silhouette, contour) is one of the most important low level image features since it is a very valuable attribute to human perception. Humans tend to perceive complex scenes as being composed of elementary objects, which can be identified by their shapes. Besides, as far as image query is concerned, shape or silhouette is simple for user to describe, either by giving example or by sketching. Moreover, most of the real objects are usually easy to detect and distinguish by means of their shapes. Shape used for recognition can be considered as a binary mask stored in a matrix, which can be represented by a specific number of points, including its interior or as a boundary (outer contour). Compact representation of shape is often known as shape descriptor. It is crucial for the recognition or classification to uniquely characterize shape and stay invariant to as many transformations as possible (i.e. translation, scaling, rotation, illumination changes, presence of noise and occlusion). These distortions are considered as differences between object under recognition and the reference object belonging to the same class, stored in a database. In real recognition tasks one has to take into consideration the following distortions divided into three main categories. The first one includes spatial transformations of an object, mainly translation, rotation in the image plane, change of scale and projection from 3D into 2D. The second category includes distortions introduced by imaging system, e.g.: variable number of points, presence of noise, discontinuity and occlusion. The third category of problems comes from contour representation and contour evaluation. The elements of the second group are the most challenging and difficult to solve. Shape descriptors can be classified in various ways. The first taxonomy is based on mentioned earlier difference between object boundary and the whole shape. The second very popular manner (as described e.g. in [2]) is based on whether the shape is represented as a whole (global approaches) or by a set of primitives (structural methods). The third one discriminates spatial and transform domains [3]. Since shape is one of the most important features in problems related to CBIR, there are many shape representations and retrieval methods known, however, most of those methods either do not represent shape in a sufficiently precise manner or are relatively challenging in matching. Among them, methods based on moments, polar coordinate system representation and histogram of distances from the centroid achieve both good representation and easy normalization [4].

Shape descriptors offer a very attractive tool for image browsing in the computer-based environment and are easily implementable using any high-level programming language. However, the main problem in porting such descriptors to pure hardware-based systems (i.e. Field Programmable Gate Array - FPGA, System on Chip - SoC, Network on Chip - NoC) is the dimensionality of feature vectors. In such case the final feature vector have to be as compact as possible, yet without losing its discriminative power, since above solutions offer a very limited resources. Several aspects of this problem have been presented in [5,6,7].

While general shape recognition and classification problems have a significant place in the scientific literature, often with reported accuracies of over 90%, the

specific topic of shape classification and retrieval by means of low-dimensional feature sets (easy to implement in hardware) is much less represented. It is probably because many well-established shape descriptors proposed in the literature give satisfactory results, yet without considering the overall computing costs. Moreover, the classification is boosted by different approaches that simply increase the dimensionality of analyzed data.

1.2 Elementary Shape Descriptors

As it was mentioned earlier, there are two main classes of shape descriptors which capture different features, namely, region-based shape representation and contour-based ones. In region based techniques, all pixels within a shape are taken into consideration to obtain final shape representation. Most of region-based methods employ different variants of moments calculation. In this article we focus on Complex Zernike Moments, as it is a well-established tool for shape retrieval. Contour-based shape representation exploit shape boundary information. Such methods can be classified into global shape descriptors , shape signatures and spectral descriptors. Although global descriptors such as area, circularity, eccentricity, axis orientation are simple to compute and also robust in representation, they can only discriminate shapes with large dissimilarities, therefore usually suitable for filtering purpose. Most shape signatures such as complex coordinates, curvature and angular representations are essentially local representations of shape features, they are sensitive to noise and not robust. In addition, shape representation using shape signatures require intensive computation during similarity calculation, due to the hard normalization of rotation invariance. As a result, these representations need often further processing.

Shape Signature. Shape signature (SSIG) is one of the most popular representations that belong to the contour-based class of descriptors. There are several variants of SSIG which employ different features. Here, we use so called centroid distance. It is easy to calculate and after some normalization can be invariant to scaling and rotation. However it should be noted, that in addition to the high matching cost, shape signatures are sensitive to noise, and slight changes in the boundary can cause large errors in matching. Hence, shape signatures should be stored in a reduced form. In this paper we calculate SSIG according to the following algorithm: (i) detect outer contour of object and store its coordinates in a polar system of coordinates (ρ, θ); (ii) find the maximal distance ρ and perform circular shift, so the θ related to this maximum occupies the first position in the vector; (iii) discard the information about θ, hence remember only distances ρ; (iv) normalize the vector to the maximal value of ρ; (v) interpolate the vector containing ρ to the final length of b elements.

Fourier Descriptor. There is a whole family of descriptors called Fourier Descriptors (FDs). Different shape signatures have been exploited to derive such descriptors. It should be noticed, that FDs derived from different signatures can

have significantly different effect on the result of retrieval. In this paper we calculate FD according to the following algorithm: (i) detect outer contour of object and store its coordinates (x, y) as complex numbers, where x coordinate is a real part, and y coordinate as imaginary part; (ii) perform Fast Fourier Transform (FFT) on these numbers; (iii) normalize FFT spectrum to its maximal magnitude; (iv) remember first b elements (related to low frequency components).

Complex Zernike Moments. The Zernike polynomials were first proposed in 1934 by Zernike [8] as tool for optical analysis. Complex Zernike moments [9] used for pattern recognition are constructed using a set of complex polynomials which form a complete orthogonal basis set defined on the unit disc. To calculate the Zernike moments, the image (or region of interest) is first mapped to the unit disc using polar coordinates, where the center of the image is the origin of such disc. Those pixels falling outside the unit disc are not considered in further calculations. The coordinates are then described by the length of the vector from the origin to the coordinate point, ρ, and the angle θ from the X axis to the vector ρ. Angular coordinate is, by convention, measured from the positive X axis in a counter clockwise direction. Translation and scale invariance can be achieved by normalizing the image using the Cartesian moments prior to calculation of the Zernike moments. Further, the absolute value of a Zernike moment is rotation invariant as reflected in the mapping of the image to the unit disc. The rotation of the shape around the unit disc is expressed as a phase change.

1.3 Fusion and Scale Normalization

The main motivation behind the presented experiments is the idea of joining known shape descriptors of low dimensionality in order to form a single compact representation, which would give much higher recognition accuracy than each of the descriptors alone. Such an approach leads to the forming of one feature vector for a single object in the database, thus the similarity evaluation means only one comparison, instead of multi-tier algorithms involving classifier cascades [10]. The problem of concatenating elementary features into single vector has also been studied by many researchers, e.g. in [11] and [12]. In the paper we present the results of investigations on fusion of above presented descriptors, namely Complex Zernike Moments, Shape Signature and Fourier Descriptors. Such descriptors were chosen in order to represent different characteristics of shapes, i.e. internal structure and external contour, respectively. They are simple to compute and give a possibility to be stored on a limited number of values. Hence, the combinations of descriptors captured both characteristics in one feature vector, which is a clear advantage over the traditional, single-characteristics approach. The fusion in our case is performed by simple concatenation of values resulted from each algorithm. The main problem with such fusion of different descriptors in a single feature vector is the difference in scales of individual representation, which strongly influences the distance calculation stage. Each elementary feature

vector is defined on different domain and the straightforward concatenation of feature vectors cannot be employed. Values in feature vector for Shape Signature depend on the size of image, hence they are normalized to the maximal value (maximal distance from the centroid). Values in Fourier Descriptor-based vector depend on the FFT spectrum and are also normalized in such way, that all resulting values are divided by maximal absolute value of the spectrum. Finally, Zernike Moments approach generates values that are much larger than in the above algorithms. They are highly correlated to the contents of image. Here, they are down-scaled by taking a logarithm of resulting absolute values of calculated moments and then normalized to the maximal value. In such way, all values are within range [0,1] and can be easily concatenated. The algorithm of fusion and normalization is presented in Fig. 1

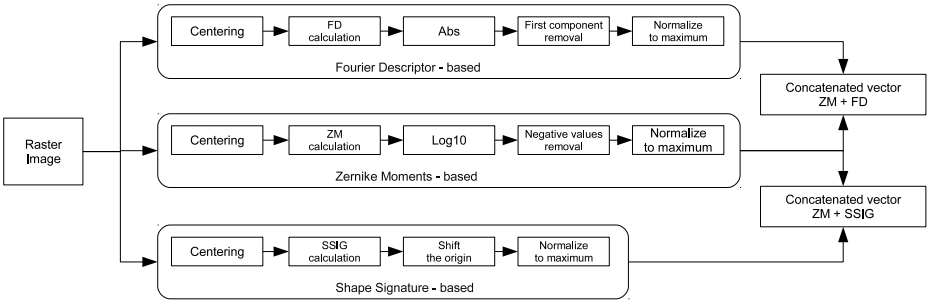


Fig. 1. Scheme of fusing features into joint descriptors

2 Experiments

We performed experiments on standard benchmark MPEG-7 Shape Database, namely Core Experiment CE-Shape-1 part B, which is used for performance evaluation of similarity-based retrieval [13]. The database contains 70 shape categories with 20 images per shape category. Each image is in a form of binary matrix, where black pixels represent background. Images have varying resolutions (minimum 128 × 128 pixels, maximum 512 × 512 pixels) and different proportions. The objects in the database belong to different categories, i.e. natural and artificial objects represented with variable accuracy, different scale, and orientation. The shapes in the database are sometimes rotated or flipped. Exemplary images from the database are presented in Fig. 2. One of crucial problems is, that objects within single class feature a very strong intra-class variation, namely although they are visually different, they can be described with a single word, e.g. 'camel' presented in last row of Fig. 2.

We performed two sets of experiments which simulate classification or retrieval tasks by means of joining above presented descriptors. In the classical approach to classification, we find the distance to the centers of classes, thus the experiments showed that the improvement can be achieved through finding the

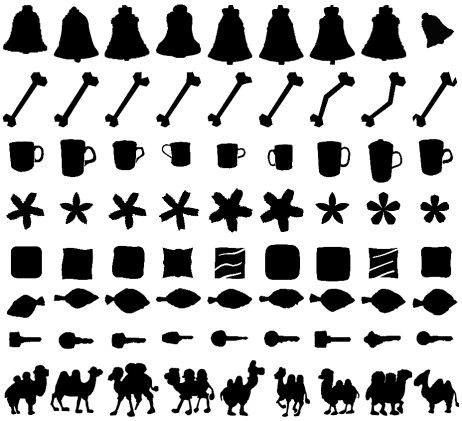


Fig. 2. Selected objects from MPEG-7 Shape Database used in the experiments

distances to several closest objects according to the k-nearest-neighbors manner. Hence, classification was performed on the basis of such rule, namely, the test image was projected into reduced feature-space and the distance to all objects in the database was calculated. The nearest neighbor pointed out the class of the analyzed object. We tested several state-of-the-art classifiers, namely Naive-Bayes (NB), MultiLayer Perceptron (MLP), Support Vectors Machine (a variant employing Sequential minimal optimization - SMO), k-Nearest Neighbors (1NN, where $k = 1$), Random Tree and Random Forest. We investigated also the influence of feature vector dimensionality (16, 32, 64, 128, 132 elements). The results of classification experiments are presented in Tab. 1. It is a mean value of True Positive Rate calculated for cross-validation with 10-fold rule. As it can be seen, single-descriptor approach gives much lower accuracy in almost any case. The best accuracy can be obtained for joint features ZM+FD and 1NN. Moreover, the dimensionality of such vector is very low (32 elements). The False Positive rate was not more than 1%, with average equal to 0.5%, which is a very good result. On the other hand, in the retrieval scenario, we want to find all objects that are similar to the test one. Here, for each test object we calculated distances to 38 nearest neighbors and checked if the proper class has its representatives. We selected 1NN as the distance measure, since it proved its performance in the earlier experiment. We counted the number of the correct answers up to the *Rank* position (see Fig. 3). Different parameters of analyzed algorithms were investigated, e.g. dimensionality of each descriptor and the size of ensemble of descriptors. We tested several variants of vector length, ranging from extremely low-dimensional vectors (9, 16 elements) through medium-low-sized (36, 49, 81) up to medium-dimensional ones (100, 121, 210 elements). During experiments with joint features we tried to create vectors with similar dimensions.

The results of retrieval showed as total accuracy are presented in Fig. 3. The plots present averaged accuracy for all dimensionalities of given feature (or feature set), together with minimal and maximal value. The first point in each

plot is related to classification accuracy ($Rank = 1$), while other points represent retrieval with proper rank. As it can be seen, the overall retrieval accuracy of single descriptors is quite high in case of ZM and SSIG and it is similar to the joint ZM+SSIG with the same vector length. Although, both single-feature and joint-features approaches give similar average results, closer analysis shows, that the ZM+FD gives better maximal accuracy for classification than any other approach, with smaller variation between minimal and maximal results. When it comes to retrieval rate, again, joint features are much better than single-feature algorithms. The increase in retrieval rate is much faster, reaching average value of over 0.99 for $Rank = 17$.

Table 1. Accuracy of different classifiers and features (single and joint) together with their dimensionalities

	SSIG				FD				ZM				ZM+SSIG	ZM+FD
Classifier	128	64	32	16	128	64	32	16	132	64	30	16	16+32	16+16
NB	71	70	68	60	67	71	74	74	55	55	54	52	74	76
MLP	75	75	73	64	67	70	72	71	44	42	40	40	80	84
SMO	79	73	68	55	51	55	54	48	53	46	37	33	77	72
1NN	83	83	80	76	58	73	79	82	81	79	72	70	82	86
Random Tree	60	60	56	55	44	53	60	64	63	63	63	65	60	70
Random Forest	77	77	72	69	64	71	74	75	76	76	74	72	76	83

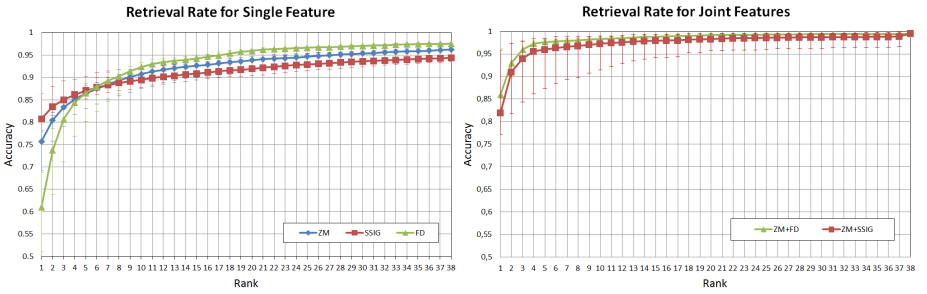


Fig. 3. Accuracy of shape retrieval for features calculated from single and joint descriptors and Euclidean distance as similarity measure

3 Summary

In the article a certain way of joining features in the task of shape classification and retrieval was presented. As it was shown, the performance is significantly better for low-dimensional features fusions than for single descriptor of the same dimensionality. With a database of 1400 images (70 classes of 20 elements each) the combination method has an accuracy of up to almost 10% better than in a case of single descriptor, in case of image retrieval, and more than 6% in case of classification.

References

1. Idirs, F., Panchanathan, S.: Review of Image and Video Indexing Techniques. Journal of Visual Communication and Image Representation 8(2), 146–166 (1997)
2. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition 37(1), 1–19 (2004)
3. Mehtre, B.M., Kankanhalli, M.S., Lee, W.F.: measures for content based image retrieval: a comparison. Information Proc. & Management 33, 319–337 (1997)
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. Kapela, R., Rybarczyk, A.: Real-time shape description system based on MPEG-7 descriptors. J. Syst. Archit. 53(9), 602–618 (2007)
6. Kapela, R., Sniatala, P., Rybarczyk, A.: Real-time visual content description system based on MPEG-7 descriptors. Multimedia Tools Appl. 53(1), 119–150 (2011)
7. Forczmański, P., Dziurzański, P.: System-Level Hardware Implementation of Simplified Low-Level Color Image Descriptor. In: Burduk, R., Jackowski, K., Kurzynski, M., Wozniak, M., Zolnierek, A. (eds.) CORES 2013. AISC, vol. 226, pp. 461–468. Springer, Heidelberg (2013)
8. Zernike, F.: Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode (Diffraction theory of the cut procedure and its improved form, the phase contrast method). Physica 1, 689–704 (1934)
9. Teague, M.R.: Image analysis via the general theory of moments. Journal of the Optical Society of America 70(8), 920–930 (1980)
10. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
11. Zhang, W., Sun, J., Tang, X.: Cat Head Detection - How to Effectively Exploit Shape and Texture Features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 802–816. Springer, Heidelberg (2008)
12. Mitsui, T., Fujiyoshi, H.: Object Detection by Joint Features based on Two-Stage Boosting. Visual Surveillance (2009)
13. Jeannin, S., Bober, M.: Description of core experiments for MPEG-7 motion/shape. Technical Report ISO/IEC JTC 1/SC 29/WG 11 MPEG99/N2690 (1999)

Accelerating the 3D Random Walker Image Segmentation Algorithm by Image Graph Reduction and GPU Computing

Jarosław Gocławski, Tomasz Węgliński, and Anna Fabijańska

Lodz University of Technology, Institute of Applied Computer Science,
18/22 Stefanowskiego Str., 90-924 Lodz, Poland
{jgoclaw,tweglinski,an_fab}@kis.p.lodz.pl

Abstract. In this paper the problem of image segmentation using the random walker algorithm was considered. When applied to the segmentation of 3D images the method requires an extreme amount of memory and time resources in order to represent the corresponding enormous image graph and to solve the resulting sparse linear system. Having in mind these limitations the optimization of the random walker approach is proposed. In particular, certain techniques for the graph size reduction and method parallelization are proposed. The results of applying the introduced improvements to the segmentation of 3D CT datasets are presented and discussed. The analysis of results shows that the modified method can be successfully applied to the segmentation of volumetric images and on a single PC provides results in a reasonable time.

1 Introduction

The main challenge of a recent image processing is the fast and efficient segmentation of large image datasets. This can be observed in particular in medical applications where a resolution of three dimensional CT and MRI body scans constantly increases.

Recently, a growing interest is attracted by an interactive graph based image segmentation algorithms such as the random walker [2]. The method has a lot advantages e.g. the ability to obey weak and discontinuous boundaries and the noise robustness. However, it is hardly ever used for segmentation of three dimensional medical data. In such a case, the usage of the method is strongly limited by the enormous size of the graph representing 3D image and the necessity of solving a huge sparse linear system. These hurdles result in the very long computation time and the usage of an extreme amount of memory resources.

Taking into account the above-mentioned limitations, this paper focuses on the problem of optimization of the random walker segmentation approach, both in terms of memory and time usage. In particular, certain techniques for the graph size reduction are proposed. Additionally, the partial parallelization of the algorithm using GPU processing and Compute Unified Device Architecture (CUDA) are introduced [6]. As a result, the method can be successfully applied

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_6

for the segmentation of volumetric medical images and provides results in a reasonable time.

2 The Idea of Random Walking for Image Segmentation

The Random Walker segmentation method [2] works on an image seen as a weighted undirected graph $G = (V, E)$ with vertices $v \in V$ corresponding to image pixels and edges $e \in E \subseteq V \times V$ spanning the pairs of vertices. Each edge $e_{ij} = (v_i, v_j)$ is assigned a non-negative weight w_{ij} related to the difference of grey level values in CT images. Six edges from any pixel p_i to its nearest neighbours are considered. The set of graph vertices is then divided into the pre-labelled nodes V_L and the unlabelled nodes V_U, where $V_M \cap V_U = \emptyset$ and $V_M \cup V_U = V$. In the Random Walker algorithm vertices from set V_U are labelled according to the probability that the random walker released from each unlabelled vertex will firstly reach a vertex labelled with a label s. Vertices are assigned labels for which the greatest probability x_U^s exists. The problem has an analytical solution in the form of a sparse, symmetric system of linear equations given in Eqn. (1)

$$\mathcal{L}_U \, x_U^s = -B f^s, \quad \mathcal{L} = \begin{bmatrix} \mathcal{L}_M & B \\ B^T & \mathcal{L}_U \end{bmatrix}, \quad \forall i, \sum_s x_i^s = 1, \tag{1}$$

where: $\mathcal{L}$ denotes the Laplacian matrix created from image graph weights, $\mathcal{L}_M$ and $\mathcal{L}_U$ are labelled and unlabelled Laplacian blocks respectively and x_U^s is the vector of probabilities, that a random walking started in each unlabelled pixel p_U will firstly reach the pixel pre-labelled by a label s.

$$f_i^s = \begin{cases} 1 & \text{if } l_i = s \\ 0 & \text{if } l_i \neq s \end{cases}, \tag{2}$$

where l_i is the label of pixel p_i. Finally the image pixel labelling takes the form:

$$\forall i \in U, \; l_i = \underset{s}{max}(x_i^s). \tag{3}$$

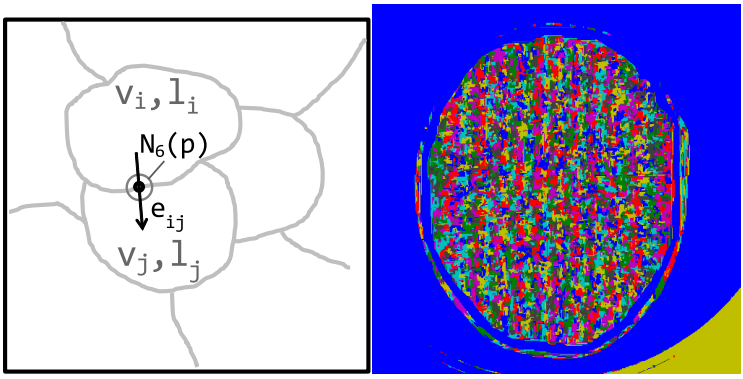
The Laplacian $\mathcal{L}_U$ size is usually huge, approximately equal to the square of the image graph order, assuming that relatively a few pixels can be labelled manually as seeds in the image space. Therefore $\mathcal{L}_U$ is stored in the host memory as a sparse matrix using the coordinate format defined in [3]. As a result, the execution of the Random Walker algorithm is very slow for a typical CT dataset.

3 Reduction of the Image Graph Size

The main idea of the introduced approach is to accelerate the classic Random Walker algorithm by the significant reduction of the number of nodes and edges in the image graph. This can be achieved by applying the idea of super-pixels

i.e. atomic, homogeneous and irregularly shaped regions within an image. The image graph is next built by assigning nodes to super-pixels rather than to single pixels and then by connecting the neighbouring super-pixels by edges. For creation of super-pixels the authors proposed an approach which uses the Minimum Spanning Tree (MST) pyramid algorithm [4].

The idea of super-pixels is explained in Fig. 1a. Super-pixels generated by the MST pyramid algorithm are shown (for an exemplary brain slice) in Fig. 1b.



(a) (b)

Fig. 1. The idea of super-pixels; (a) the illustration of an image graph with irregular regions and edges between region neighbours; p - edge pixel of the image graph vertex v_i; $N_B(p)$ - the neighbourhood of the pixel p; l_i, l_j - the labels of the example regions v_i, v_j; e_{ij} - the edge between v_i and v_j; (b) visualisation of super-pixels provided by MST pyramid algorithm

The introduced method for super-pixel creation in three-dimensional images is outlined in Algorithm 2.. The method creates a series of image region partitions, each of them using the parallel algorithm of MST building formulated by Boruvka [9]. At the k-th pyramid level the smallest weight edges $ME^{(k)}(i)$ are collected around the edge connected components $CC^{(k)}(i)$ in step 3. The edges to be contracted to the components are filtered in step 4 by Felzenschwalb-Huttenlocher comparison predicate [1]. For the purpose of image graph reduction the algorithm is additionally stopped at the pyramid level, when the mean count of super-pixels $CV^{(k)}$ exceeds the assumed limit value CV_{min}.

4 Image Graph Creation Based on Super-Pixels

To perform the Random Walker image segmentation efficiently, the image graph should be built at minimum time cost. Therefore, in the introduced approach the parallel computing is applied in order to determine coordinates of graph

Algorithm 2. MST pyramid segmentation for image graph reduction

Input: $G^{(0)}(CC^{(0)}, E)$ – image graph
Output: $G^{(0)}, G^{(1)}, \ldots, G^{(k)}$ – graphs at each level of the pyramid
1: k=0
▷ $CC^{(k)}(i)$ – i-th region of $G^{(k)}$
2: **repeat**
3: $ME^{(k)} \leftarrow$ the smallest edges around each $CC^{(k)}(i)$
4: $CE^{(k)} \leftarrow ME^{(k)}$ filtered by the Felzenszwalb predicate
5: $CV^{(k)} \leftarrow$ mean size of the regions $CC^{(k)}(i)$
6: $G^{(k+1)} \leftarrow \{G^{(k)} \cup CE^{(k)}\}$
7: $k \leftarrow k + 1$
8: **until** $CV^{(k)} > CV_{lim} \vee G^{(k)} = G^{(k-1)}$

edges and the corresponding edge weights. These are computed with regard to properties of super-pixels.

The procedure of image graph creation with regard to super-pixels is presented in Algorithm 3.. It requires a kernel counting of the predicted edges with the condition shown in step 2. To obtain the edge array E the count texture buffer is exclusively scanned using CUDA *thrust* template library [7].

Selected edges must be sorted by *thrust* function to eliminate possible duplicated items (step 4). After intensity normalisation, edge weights are transformed by exponential weighting function. For millions of super-pixels in CT images the times of data transfer between host an device memories can be approximately neglected.

Algorithm 3. The GPU based algorithm of building the reduced image graph from irregular super-pixels in three dimensions

Input: L – super-pixel label image, I – label grey levels
Output: E – super-pixel edges, W – super-pixel weights

1: copy L, I images to device memory
2: count & thrust scan boundary edges, while $L(p) < L(q),\ q \in N_6(p)$
3: $E \leftarrow$ store edge coordinates
4: thrust sort E, remove item duplications
5: normalise image grey-levels I
6: map exponentialy grey-levels to edge weights
7: copy edges and weights to host memory

5 Experimental Results

This section presents the results of applying the considered graph size reduction method to the Random Walker segmentation of example CT brain datasets. The tests were performed on a single PC computer equipped with a 6-core,

3.5 GHZ *Intel I7* processor and using *Windows 7 64 bit* operating system with 32 GB RAM memory. The program codes were included in C++ MEX files called from *MATLAB 2013* environment. GPU computing was performed using *Nvidia GeForce GTX Titan* graphic card with 6144 MB of device memory.

The proposed approach was tested on 10 CT brain datasets. They consisted of $115 - 336$ slices, 512×512 pixels each. Several example slices from these datasets are shown in Fig. 2. The corresponding results of image segmentation are marked with grey colour. Additionally, under each slice 3D visualization of the segmentation result is given.

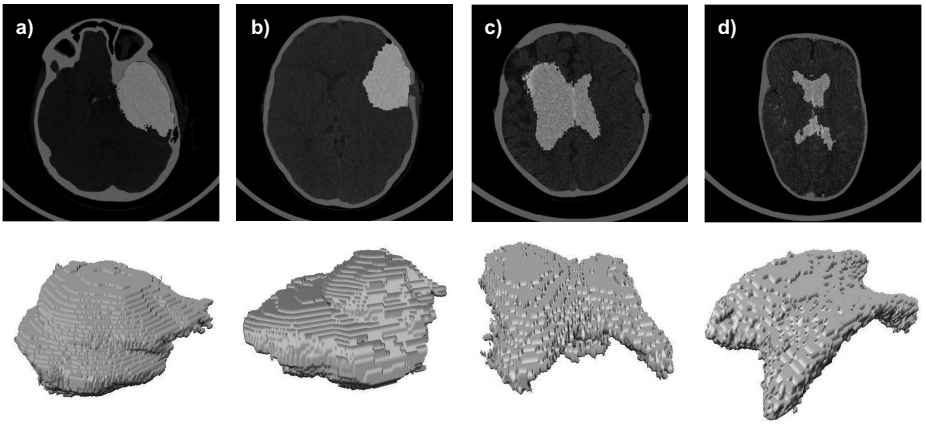


Fig. 2. Example slices of brain CT images with different lesions detected by the proposed Random Walker algorithm and the corresponding 3D visualization of the segmentation result. The diseased regions are exposed as grey regions.

The level of graph size reduction and resource usage by the introduced approach are summarized in Table 1. In particular, the case ID is given in the first column. The corresponding number of slices is given in the second column. This is followed by the time of method execution, which includes: time of image division into super-pixels (series: *Pyramid*), time of graph creation (series: *Graph*) and the total time of method execution including Random Walker segmentation (series: *Total*) given in columns number 3, 4 and 6 respectively. Two last columns consider memory usage by GPU (series: *Device*) and host RAM memory (series: *Host*).

The results presented in Table 1 show that the proposed approach significantly reduces the size of image graph. In particular, the size of the graph build upon super-pixels is on average equal to 1.25% of the original graph (i.e. the one build based upon single pixels). The graph size reduction is then on average $98,75\%$. The total time of algorithm execution vary from $17s$ to $65s$. This time is acceptable, especially taking into account that the running time of the original method is extremely long and sometimes it is impossible to run the original Random Walker on a single PC due to lack of host memory resources.

Table 1. Execution time and memory workload for the accelerated Random Walker segmentation algorithm. A common size of CT image slices: $512 \times 512 px$, Graph Reduct. - the percentage of graph size reduction

Case ID [-]	Num Slices [-]	Graph Reduct. [%]	Time Pyramid [sec.]	Time Graph [sec.]	Time Total [sec.]	Memory Device [MB]	Memory Host [MB]
1	115	98.51	15.65	0.28	20.28	1320	2172
2	336	98.89	46.64	0.48	64.65	2284	5880
3	172	98.72	22.27	0.33	31.50	1320	3376
4	200	98.80	26.61	0.36	39.77	1469	4053
5	110	98.91	14.23	0.27	17.35	1320	1870
6	207	98.60	27.60	0.36	44.00	1497	4474
7	200	98.80	27.04	0.36	40.13	1471	4133
8	204	98.63	27.12	0.37	41.16	1320	4315
9	156	98.90	20.35	0.32	26.38	1320	2700
10	267	98.78	33.95	0.41	53.35	1735	5362

Division of an image into super-pixels (by creation of MST pyramid) is an important step which takes on average $\approx 70\%$ of the total execution time. However, this step ensures a considerable reduction of the number of graph vertices and the resulting graph size.

In the proposed approach the Eqn. (1) was solved by calling the built-in $MATLAB$ solver for symmetric and sparse Laplacian matrix.

The comparison of the introduced approach (series: *Fast RW*) with the original Random Walker approach (series: *Raw RW*) is given in Table 2. The comparison considers both the usage of time and memory resources as well as the image segmentation accuracy. Because it was impossible to run the original Random Walker approach on the whole input CT datasets, the comparison was performed using the data subset of 10 consecutive slices from each dataset.

Results presented in Table 2 clearly show, that the execution time of the accelerated algorithm is on average 30 times shorter than in the case of the original Random Walker approach. Additionally, the host memory usage by the new approach exceeds merely 5.4% of the original method.

The F-score accuracy measure [5] listed in the last column of Table 2 fits in the range $0.87 \div 0.98$ (with an average value equal 0.93) indicating good compatibility between segmentation results provided by both regarded segmentation methods.

The accelerated method still uses the standard random walking parameter β applied to proper mapping of graph edge weights. In the described tests β was equal to 600. The new method also introduces an additional parameter $CV^{(k)}$ which limits regions size and stops the MST pyramid growth. The latter parameter was set experimentally to $CV^{(k)} = 1000\, px$. However, it should be underlined, that the algorithm behaviour is insensitive to small changes of this parameter.

Table 2. Resource usage and relative accuracy for the original and the accelerated Random Walker segmentation algorithm. CT image size: $512 \times 512 \times 10px$, F-score: accuracy measure,Graph Reduct.: the percentage of graph size reduction

Case ID [-]	Graph Reduct. [%]	Raw RW Time [sec.]	Raw RW Memory [MB]	Fast RW Time [sec.]	Fast RW Memory [MB]	F-score [-]
1	98.45	62.55	12916	2.04	725	0.98
2	98.37	71.70	13694	2.05	706	0.96
3	98.55	66.35	13471	1.96	719	0.94
4	98.43	65.40	13197	2.03	680	0.96
5	98.68	69.16	13884	2.06	733	0.94
6	98.43	74.32	13439	2.24	731	0.93
7	98.43	69.32	13608	1.98	735	0.97
8	98.32	71.29	13423	2.06	736	0.97
9	98.64	74.83	13940	1.95	742	0.92
10	98.95	67.96	13555	1.99	733	0.87

6 Conclusions

This paper considered the problem of acceleration and optimization of the Random Walker segmentation algorithm. The advantages of this method, like robustness to the weak or vanishing edges, predispose it to the segmentation of medical images. However, due to the very long execution time and the extreme memory workload for medical datasets (containing billions of pixels), the method cannot be directly applied for image processing.

Improvements proposed in this paper eliminate the main hurdles of the Random Walker segmentation approach. This goal is obtained by both: applying the idea of super-pixels for image graph creation and using GPU computing. The proposed algorithm reduces an image neighbourhood graph about 80 times for full size images and accelerates the segmentation above 30 times for the blocks of 10 slices. To make the full size comparison appropriate computer grid will be required. In the future also other pre-segmentations like parallelised region growing will be tested to improve the method acceleration.

The proposed improvements enable applying the method to segmentation of large three dimensional CT images using a single PC. Additionally, it provides results in a reasonable time. As a result, the number of possible applications of the Random Walker approach is increased.

Acknowledgements. This research was funded by the Ministry of Science and Higher Education of Poland from founds for science in years 2013-2015 in a framework of Iuventus Plus Programme (project no. IP 2012 011272).

References

1. Felzenschwalb, P., Huttenlocher, F., Efficient, D.P.: graph-based image segmentation. International Journal of Computer Vision 59, 1–25 (2004)
2. Grady, L.: Random Walks for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11), 1–17 (2006)
3. Intel Sparse matrix storage formats (2013), http://software.intel.com/sites/products/documentation/hpc/mkl/mklman/GUID-9FCEB1C4-670D-4738-81D2-F378013412B0.htm
4. Ion, A., Kropatsch, W.G., van Haxhimusa, Y.: Considerations Regarding the Minimum Spanning Tree Pyramid Segmentation Method. In: Yeung, D.-Y., Kwok, J.T., Fred, A., Roli, F., de Ridder, D. (eds.) SSPR&SPR 2006. LNCS, vol. 4109, pp. 182–190. Springer, Heidelberg (2006)
5. Labatut, V., Cherifi, H.: Accuracy Measures for the Comparison of Classifiers (2012), http://arxiv.org/ftp/arxiv/papers/1207/1207.3790.pdf
6. Nvidia Corporation Cuda Toolkit Documentation (2013), http://docs.nvidia.com/cuda/index.html
7. Nvidia Corporation Thrust Quick Start Guide (2013), http://docs.nvidia.com/cuda/pdf/Thrust_Quick_Start_Guide.pdf
8. Pratt, W.K.: Digital Image Processing, 4th edn. John Wiley & Sons, Inc., Los Altos (2007)
9. Vineet, V., Harish, P., Suryakant, P.: Fast Minimum Spanning Tree for Large Graphs on the GPU. In: Proceedings of the Conference on High Performance Graphics, pp. 167–171 (2009)

Computed Tomography Images Denoising with Markov Random Field Model Parametrized by Prewitt Mask

Michał Knas and Robert Cierniak

Institute of Computational Intelligence,
Armii Krajowej 36, 42-200 Czestochowa, Poland
{michal.knas,robert.cierniak}@iisi.pcz.pl

Abstract. A denoising algorithm for computed tomography images is proposed. The presented method of noise reduction uses Markov Random Field (MRF) model, Gaussian filter and adaptive Prewitt Mask, what gives better results than standard approach of using only the MRF. This implementation on Compute Unified Device Architecture is made, what makes this computationally complex denoising method faster.

1 Introduction

Computed tomography images are very useful in medicine as a diagnostic tool. This technique gives ability to view organs and tissues not only in the coronal, sagittal or axial planes, but also in other, different, not orthogonal planes. With appropriate algorithms we can also better detect the edges of the tissue and than easly render 3D presentation of human body organs, which is our main objective.

However, CT also has drawbacks. Besides the very high prices of equipment, its adverse effect is unquestionably the possibility of damage the DNA that may lead to cancer. One possible way to decrease the risk of complications after CT scans is reduction of radiation dose. Although, the data collected from the sensors becomes then more noisier. Considering it that the images are stored in digital form, beyond the improvement of equipment, we can also easily approach to the problem of noise from the mathematical point of view [1].

To get good medical diagnostic material, it is necessary to conduct the process of denoising. There are plenty of methods to accomplish that, for example linear smoothing filters, anisotropic diffusion, non-local means or nonlinear filter [2]. In our case, the priority is to choose a denoising mechanism that does not result in loss of image detail, especially near edges of different tissues. Because of its enormous diversity and complex structure of the human body, we decided to create an algorithm based on the MRF model, which takes into consideration the local characteristics of the image and makes correct classification of large random values [3,4]. This method gives very good results in segmentation and denoising but it is computationally prohibitive and also it is not always easy to

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_7

set the corresponding parameters. In the next section we describe how to solve these problems.

2 Proposed Algorithm

In our algorithm, we can distinguish three essential components: Prewitt mask, Gaussian filter and MRF. All of them are well known in the image processing, but they have been appropriately modified or adapted for the needs of our task.

2.1 Prewitt Operator

First of all, in our algorithm, we use the Prewitt operator for edge detection. We decided to use this method because of its simplicity. For the better results and the sharper edges, we decided also to adaptive approach [5]. However, the convolution function created by us, rotates mask by any angle and choose a solution where the matching of operator and edge is the best. These are the highest values at a given point, and they are written into the resulting matrix M. This rotation with Prewitt mask is very simple and shows by Equation (1) and Fig. 1.

$$z = -x\sin(\alpha\pi/180) + y\cos(\alpha\pi/180) \tag{1}$$

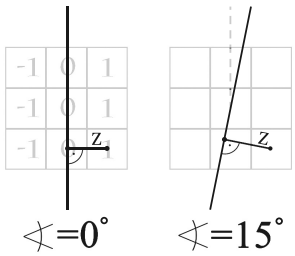


Fig. 1. Rotation of Prewitt Mask

2.2 Gaussian Filter

The next step is a preliminary denoising, where we use the gaussian filter. However, to prevent the loss of tissue boundaries, each data point is smoothed using a two-dimensional Gaussian function (2) parameterized by a previously calculated matrix M. Only if the average value of the eight nearest neighbors pixels is not significantly different from the average nearest twenty-four neighbor pixels.

$$f(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\left(\frac{(x-x_o)^2}{2\sigma_x^2} + \frac{(y-y_o)^2}{2\sigma_y^2}\right)\right), \tag{2}$$

where:

$$\sigma = \frac{k(max(M) - M[i,j])}{max(M)} \tag{3}$$

and k is determined experimentally.

2.3 Markov Random Field Model

Last part of our algorithm is based on Markov Random Field model. It means that, first of all, we have to create a set of random values. In our case it is simply one per pixel. Our matrix consists $n \times m$ random values where n and m are width and height of CT scan. In the next step, we create a graph. We decided to create it with standard nearest-neighbor dependence. Fig. 2 shows these cliques.

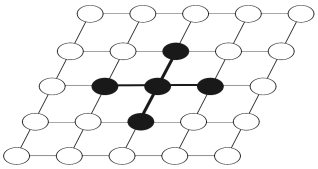


Fig. 2. Cliques of Markov Random Field model in our algorithm

The third step is to create ranges of random values. To reduce the computational complexity and at the same time taking into account the importance of the construction of the Hounsfield scale, and adequate accuracy of the calculations, we automatically created thousands of ranges between the lowest and highest value in the clique.

The last thing to do, is to create energy functions and optimization. Our equation, in line with the clique, is the sum of single-site potential and pairwise potential, which shows following equation:

$$E(x) = \frac{1}{2\sigma} \sum_{i \in C} \psi_i(x) + \sum_{i,j \in C} \psi_i(x - x_j). \tag{4}$$

$\psi(\cdot)$ is the potential function and σ is dependent on the previously calculated matrix M, what we can see below:

$$\sigma = p(max(M) + 1 - M[i,j]), \tag{5}$$

where p expresses the expected variance of Gaussian noise.

To improve the results, we minimized the total energy by finding the smallest potential in the clique. We used Iterate Conditional Mode. This simple algorithm is always convergent, so it is easy to define exit criteria.

3 Experimental Results

We carried out some computer simulations to verify whether our modifications will give better results than standard approaches.

To create experimental environment we used Microsoft Visual Studio 2008 with CUDAfy library and NVIDIA CUDA Toolkit 5.0 to achieve better performance and real time results. Tests were performed on NVIDIA GTX 680 and on Intel Core2Quad Q8400. The testing platform was Windows 8. It was necessary to use the GPU. In effect, despite the very high computational complexity, we received the results in less than one second, even at ten iterations.

We tested our algorithm on three data sets. Two of them were phantoms with resolution of 1024 *times* 1024, and one was a real image stored in the 16-bit DICOM format with resolution 512 *times* 512. For each of these images, we added noise with normal distribution, where expected value equals zero and variance equals 0.02 for phantoms and 0.05 for real scan.

The first phase of testing was focused on the edge detection and showed that the twelvefold rotation of 3 *times* 3 Prewiit operator about 30° gives the best results, what we can see in Table 1.

Table 1. MSE of edge detection on Phantom image with 1024 resolution

	15°	20°	30°	45°	60°
3×3 mask	3,22e-03	3,15e-03	2,91e-03	3,20e-03	5,17e-03
5×5 mask	36,46e-03	35,73e-03	33,44e-03	47,42e-03	142,35e-03

Then we checked the MSE of images after denoising by the standard gaussian filter, the MRF and by our preliminary denoising which uses the Prewitt operator. Table 2 and Fig. 3 shows these results.

Table 2. MSE of image denoising

	Noised	Gaussian Filter	GF and Prewitt	MRF
Phantom	39,795e-05	61,301e-05	2,623e-05	3,966e-05
Forbild	39,793e-05	84,793e-05	5,634e-05	6,548e-05
DICOM	247,763e-05	108,556e-05	55,903e-05	69,129e-05

After this tests, we decided to use MRF over denoised image with Prewitt operator what gave the best results as it shows Fig. 4.

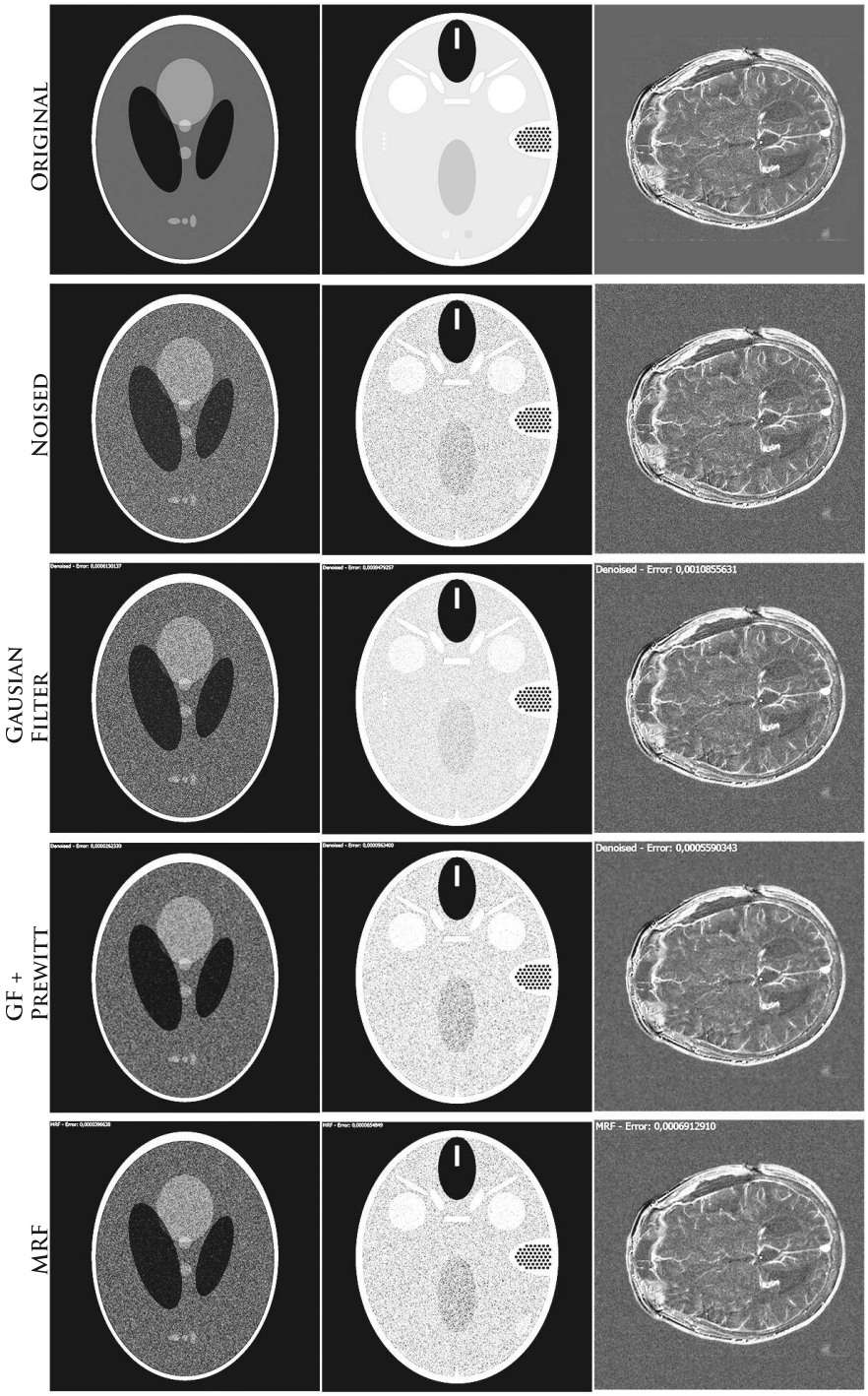


Fig. 3. Original, noised and denoised images

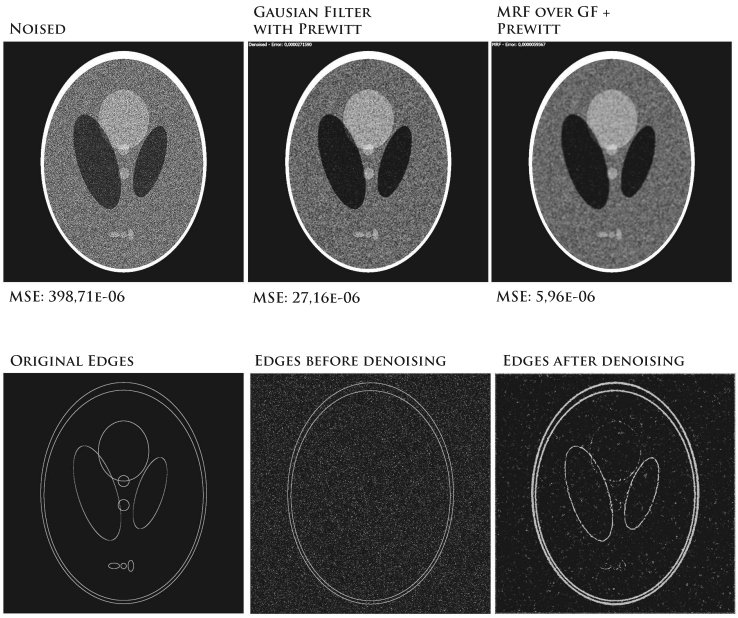


Fig. 4. MRF over denoised image with Prewitt operator

4 Conclusion

Tests performed by us have shown that our modifications yield satisfyingly good results. In all data we significantly reduced the noise without loosing the sharpness and important details.

Our algorithm is computationally complex, but using GPUs allows the fast and convenient denoising.

The effectiveness of our solution depends on the choosen parameters. To facilitate that, we were analysed the histograms of examined scans.

The most important for us is the fact that the level of edge detection significantly increased when we were using our algorithm.

References

1. Cierniak, R.: X-Ray Computed Tomography in Biomedical Engineering. Springer, London (2011)
2. Weickert, J.: Anisotropic diffusion in image processing, vol. 1. Teubner, Stuttgart (1998)
3. Panjwani, D.K., Healey, G.: Markov random field mod-els for unsupervised segmentation of textured color images. IEEE Trans. on Pattern Analysis and Machine Intelligence 17(10), 939–954 (1995)
4. Yang, X.Y., Liu, J.: Unsupervised texture segmentation with one-stepmean shift and boundary Markov random fields. Pattern Recognition Letters 22(10), 1073–1081 (2001)
5. Kaushal, M., Singh, A., Singh, B., Kaushal, M.: Adaptive Thresholding for Edge Detection in Gray Scale Images. International Journal of Engineering Science and Technology 2 (2010)

Neural Video Compression Algorithm

Michał Knop and Piotr Dobosz

Institute of Computational Intelligence, Czestochowa University of Technology,
Armii Krajowej 36, 42-200 Czestochowa, Poland
{michal.knop,piotr.dobosz}@iisi.pcz.pl

Abstract. In this paper, we present and discuss experimental results of a new concept of algorithm for video compression. It is named Predictive Vector Quantization (PVQ) and incorporates competitive neural network quantizer and neural network predictor. It is important for the image compression based on this approach to correctly detect scene changes in order to improve performance of the algorithm. We describe the scene detection algorithm based on the image entropy and discuss its effectiveness.

1 Introduction

Multimedia data transmission is widely spread nowadays. Most of the applications require effective data compression in order to decrese the required bandwidth or storage space. Various techniques of coding of the data achieve this goal by reducing data redundancy. In most of the algorithms and codecs a spatial compensation of images as well as movement compensation in time is used. Video compression codecs can be found in such applications as:

1. various video services over the satellite, cable, and land based transmission channels (e.g., using H.222.0 / MPEG-2 systems);
2. by wire and wireless real-time video conference services (e.g., using H.32x or Session Initiation Protocol (SIP) [1]);
3. Internet or local area network (LAN) video streaming [2];
4. storage formats (e.g., digital versatile disk (DVD), digital camcorders, and personal video recorders) [3].

Currently, many image and video compression standards are used. The most popular are JPEG and MPEG. They differ in the level of compression as well as application. JPEG and JPEG2000 standards are used for image compression with an adjustable compression rate. There is a whole family of international compression standards of audiovisual data combined in the MPEG standard, which is described in more details in literature [4]. The best known members are MPEG-1, MPEG-2, and MPEG-4. We used a PVQ (Predictive Vector Quantization) algorithm in our work to compress a video sequence. It combines a VQ (Vector Quantization) [5,6] and DPCM (Differential Pulse Code Modulation). More information on the techniques can be found in sources [7,8,9]. To detect a scene change we used method based on the entropy. Then we can change necessary parameters of the predictor and the codebook.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_8

2 Video Compression Algorithm

The design of the compression algorithm described here is based on the existing algorithm described in [7,8,9]. We chose this algorithm because thanks to neural network it exhibits better adjustment parameters to a frame and gives better image quality after compression. The extension includes a scene change detection algorithm, which is based on the entropy of frames. The diagram below Fig. 1. shows the proposed algorithm.

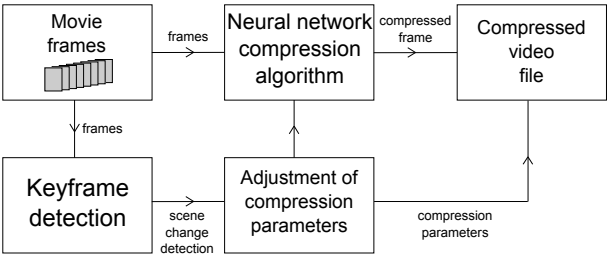


Fig. 1. Video compression algorithm

2.1 Neuronal Image Compression

Fig. 1 [9] shows the architecture of the PVQ. It combines the Huffman coding with a vector extension of the scalar differential pulse code modulation scheme [7,8]. The block diagram of the PVQ algorithm consists of the encoder and decoder, each containing: an identical neural-predictor, a codebook, a neural vector quantizer, the Huffman coder.

The successive input vectors $\mathbf{V}(t)$ are introduced to the encoder. The difference $\mathbf{E}(t) = [e_1(t), e_2(t), ..., e_q(t)]^T$ given by the equation

$$\mathbf{E}(t) = \mathbf{V}(t) - \overline{\mathbf{V}}(t) \tag{1}$$

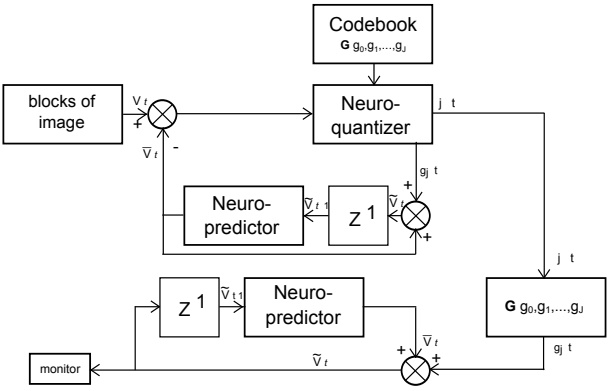


Fig. 2. The architecture of the image compression algorithm

is formed, where: $\overline{\mathbf{V}}(t) = [\overline{v}_1(t), \overline{v}_2(t), ..., \overline{v}_q(t)]^T$ is the predictor of $\mathbf{V}(t)$. Statistics shows that fewer quantization bits are required by the difference $\mathbf{E}(t)$ than the original subimage $\mathbf{V}(t)$. The next step is vector quantization of $\mathbf{E}(t)$ using the set of reproduction vectors $\mathbf{G} = [\mathbf{g}_0, \mathbf{g}_1, ..., \mathbf{g}_J]$ (codebook), where $\mathbf{g}_j = [g_{1j}, g_{2j}, ..., g_{qj}]^T$ (codewords). For every q-dimensional difference vector $\mathbf{E}(t)$, the distortion (usually the mean square error) between $\mathbf{E}(t)$ and every codeword $\mathbf{g}_j$, $j = 0, 1, ..., J-1$ is computed. The codeword $\mathbf{g}_{j^0}(t)$ is selected as the representation vector for $\mathbf{E}(t)$ if

$$d_{j^0} = \min_{0 \le j \le J} d_j, \tag{2}$$

a measure d in expression (2)we can take e.g. the Euclidean distance. Observe that by adding the prediction vector $\overline{\mathbf{V}}(t)$ to the quantized difference vector $\mathbf{g}_{j^0}(t)$ we get the reconstructed approximation $\widetilde{\mathbf{V}}(t)$ of the original input vector $\mathbf{V}(t)$, i.e.

$$\widetilde{\mathbf{V}}(t) = \overline{\mathbf{V}}(t) + \mathbf{g}_{j^0}(t). \tag{3}$$

The prediction vector $\overline{\mathbf{V}}(t)$ of the input vector $\mathbf{V}(t)$ is made from past observation of reconstructed vector $\widetilde{\mathbf{V}}(t-1)$. In our algorithm, the predictor is specifically designed for this purpose as a nonlinear neural network. As the last stage, the set of the $j^0(t)$ is coded by the Huffman coder. The codebook of the Huffman coder is designed using a set of counters f_j which count how frequently given label $j^0(t)$ occurs after presentation of all vectors $\mathbf{V}(t)$. The appropriate codewords $h^0(t)$ from the Huffman codebook are broadcasted via the transmission channel to the decoder, where they are decoded and the reconstructed vector $\widetilde{\mathbf{V}}(t)$ is formed in the same manner as in the encoder (see formula (3)).

2.2 Scene Detection

Among many other approaches [10,11,12], the methods based on the entropy are worth to take into account. We were used the phenomenon of the entropy to determine the complexity of information stored in each film frames. The general formula for the entropy on the greyscale video materials looks like:

$$E_j = -\sum_{i=0}^{N} p_i log(p_i), \tag{4}$$

where E_j is a characteristic value calculated from current frame, p_i is the next pixel from this frame, and N is a number of whole pixels in current image. Since we have been used *HSV* color gamut, we must calculate the value of the entropy for each coefficient in palette. The individual components of the *HSV* model aren't equal. The hue *(H)* has the biggest importance in the process of determining the coefficient. Even the slightest change of its value should lead to detecting scene changes. A bit less importance has the color saturation *(S)*, which even larger changes are for us fully acceptable. The value of the color *(V)*

has for us the lowest importance. Hence the final value of entropy for each frames in the *HSV* color space will be calculated from the formula:

$$F_n(E) = a \cdot E_j(H) + b \cdot E_j(S) + c \cdot E_j(V) \tag{5}$$

Parameters of a, b and c corresponds to the weighting factors for the entropy component of *HSV*. For our experiment, the values are successively set to values 0.9, 0.3 and 0.1 [13]. To detect new keyframe we compare current keyframe with subsequent frames. The following formula describes used compare operation:

$$E_{j,k} = \frac{|E_j - E_k|}{E_k}, \tag{6}$$

where j is a number of our current frame and k is a current key frame. The first key frame is always the first image taken at start of the film. The next key frames are being calculated based on the comparison result (see formula 6). The threshold value for determining a keyframe was selected experimentally and is set to be -0.2. If the entropy value is lower than the assumed threshold, the algorithm determines a new key frame.

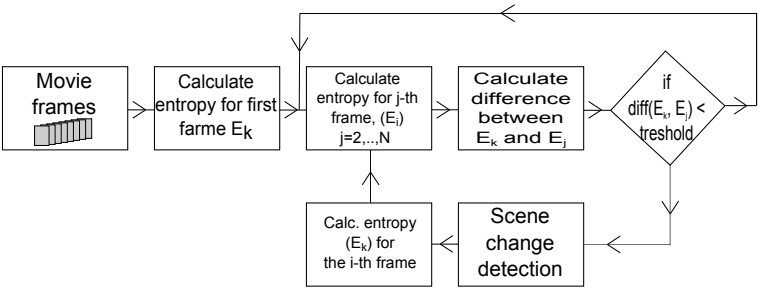


Fig. 3. Scene change detection algorithm

3 Experimental Result

The effectiveness of the algorithm was tested on a set of uncompressed frame captured by a digital USB camera of a 640×480 resolution with 256 levels of grey. Four tests were conducted. In the first and second test, we used the frames within a single scene (Fig. 4). In the first test, the frames were compressed creating separate codebook and the predictor for each frames (Fig. 5). For the second test we used the same codebook and the predictor for all frames(Fig. 6).

A transit frames between scenes were chosen for the third and fourth tests based on the scene change detection algorithm (Fig. 7). In this algorithm each frame is compare with the keyframe. When the new scene is detected the algorithm marks a new keyframe (Fig. 8). Thanks to the training of predictor and codebook for the keyframe, algorithm adapts better compression parameters to the set of frames and improves the image quality after decompression (Fig. 9).

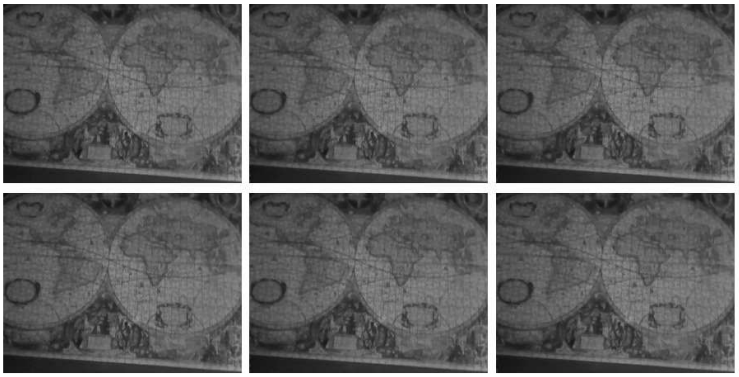


Fig. 4. a)original sequence b)compressed sequence test 1 c)compressed sequence test 2

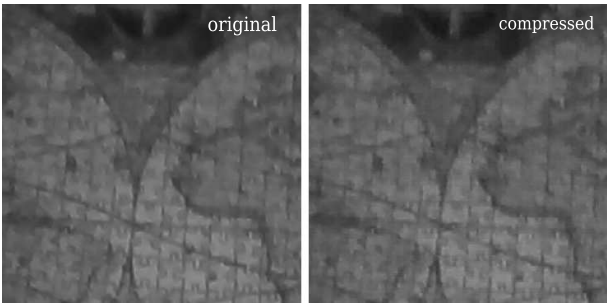


Fig. 5. Difference between frames in test 1

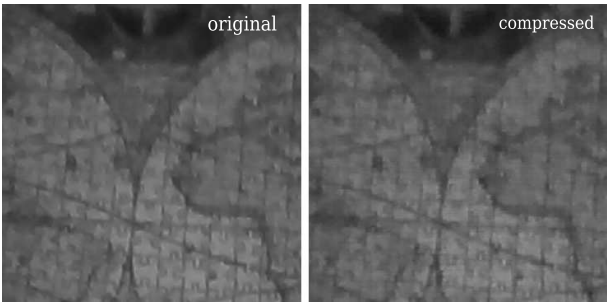


Fig. 6. Difference between frames in test 2

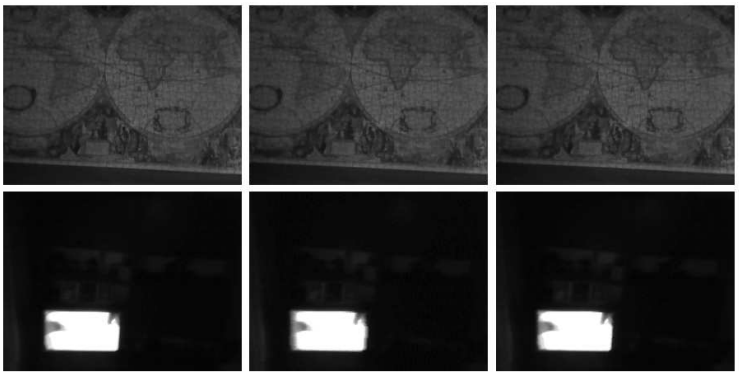


Fig. 7. a)original sequence b)compressed sequence test 3 c)compressed sequence test 4

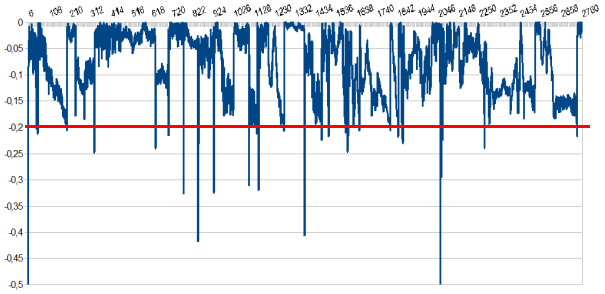


Fig. 8. Scene change detection

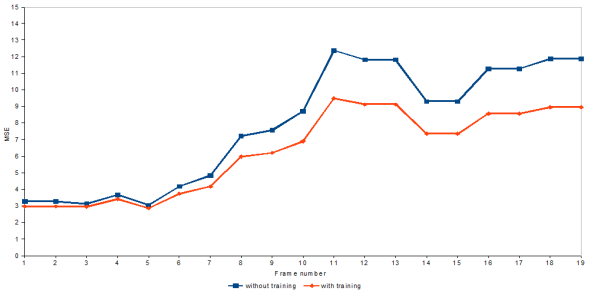


Fig. 9. MSE change

In test 3 the same codebook and predictor were used before and after the scene transit. As the results show, this approach is insufficient in case of a major scene change (Fig. 10). For the fourth test, the scene transition was detected and separate codebooks and predictors were created for frames before and after the scene transition (Fig. 11).

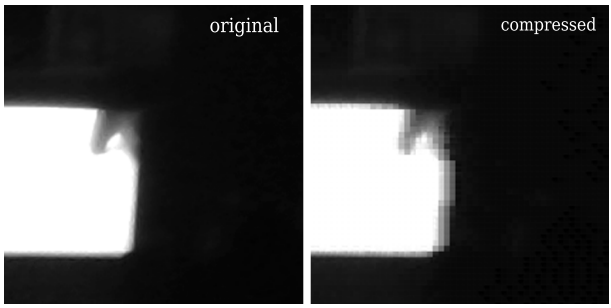


Fig. 10. Difference between frames in test 3

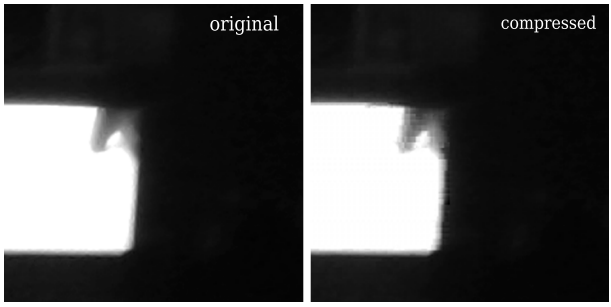


Fig. 11. Difference between frames in test 4

4 Conclusions

The tests show that the scene change detection algorithm is especially useful for the presented compression algorithm. It is apparent that without the scene detection video sequence compressed by our algorithm would exhibit a poor quality of frames after the scene transition. On the other hand, the amount of data resulting from including the compression parameters for every frame would greatly impact the output file's size.

References

1. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E.: SIP: Session Initiation Protocol. Internet Eng. Task Force (IETF). Request for Comments (RFC), 3261 (2002)

2. Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V.: RTP: A transport protocol for real-time applications. Internet Eng. Task Force (IETF). Request for Comments (RFC), 1889 (1996)
3. Sullivan, G.J., Wiegand, T.: Video Compression—From Concepts to the H.264/AVC Standard. Proceedings of the IEEE 93, 18–31 (2005)
4. Clarke, R.J.: Digital compression of still images and video. Academic Press, London (1995)
5. Gray, R.: Vector quantization. IEEE ASSP Magazine 1, 4–29 (1984)
6. Gersho, A., Gray, R.M.: Vector quantization and signal compression. Kluwer Academic Publishers (1992)
7. Rutkowski, L., Cierniak, R.: Image compression by competitive learning neural networks and predictive vector quantization. Applied Mathematics and Computer Science 6, 706–711 (1996)
8. Cierniak, R., Rutkowski, L.: On image compression by competitive neural networks and optimal linear predictors. Signal Processing: Image Communication 15, 559–565 (2000)
9. Cierniak, R.: An Image Compression Algorithm Based on Neural Networks. In: Rutkowski, L., Siekmann, J.H., Tadeusiewicz, R., Zadeh, L.A. (eds.) ICAISC 2004. LNCS (LNAI), vol. 3070, pp. 706–711. Springer, Heidelberg (2004)
10. Chen, L., Feris, R., Turk, M.: Efficient partial shape matching using Smith-Waterman algorithm. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, pp. 1–6 (2008)
11. Adhikari, P., Gargote, N., Digge, J., Hogade, B.G.: Abrupt Scene Change Detection. World Academy of Science, Engineering and Technology. International Science Index 18, 687–692 (2008)
12. Seeling, P.: Scene Change Detection for Uncompressed Video. In: Technological Developments in Education and Automation, pp. 11–14 (2010)
13. Qu, Z., Lin, L., Gao, T., Wang, Y.: An Improved Keyframe Extraction Method Based on HSV Colour Space. Journal of Software 8, 1751–1758 (2013)

Discovering Important Regions of Cytological Slides Using Classification Tree*

Marek Kowal[1], Andrzej Marciniak[1],
Roman Monczak[2], and Andrzej Obuchowicz[1]

[1] Univeristy of Zielona Góra,
Institute of Control and Computation Engineering, Poland
{M.Kowal,A.Marciniak,A.Obuchowicz}@issi.uz.zgora.pl
[2] Department of Pathomorphology, Regional Hospital in Zielona Góra, Poland
R.Monczak@issi.uz.zgora.pl

Abstract. Modern digital microscopy systems allow imaging of biological material with very high accuracy. Paradoxically, this gives rise to many problems because huge amounts of raw data significantly increase the time required by specialist to analyze them. As a result, we obtain a time-consuming diagnostic process and reduction of the number of patients being diagnosed. The paper presents a method of discovering regions of the cytological image, which are essential to correct diagnosis. The purpose of this method is to help pathologists by indicating regions of the image that should be analyzed first. Moreover, method can be used to explore new, previously unknown features discriminating benign from malignant lesions. Multi-level image thresholding is responsible for image segmentation and is the core of the proposed system. Thresholds are evaluated by predictive accuracy on testing dataset. Honey Bee Mating Optimization (HBMO) algorithm is applied to find the optimal threshold set. The developed method was successfully applied to analyze cytological images of biological material collected from a breast tumor.

1 Introduction

Nowadays whole slide digital imaging is becoming a standard in cytological examinations. Glass slides are scanned, converted to digital slides and can be viewed on computer monitor. The scanning accuracy is usually very high, and the output slide has a large size. Visual analysis of the entire slide can then be very time-consuming. To overcome this problem, numerous automatic segmentation algorithms was proposed to extract features of cells or nuclei from cytological images. Unfortunately, it still remains a big challenge due to noise, cell and nuclei overlapping or high variation of image parameters. Most of the approaches lack of generality, and give expected results only for specific images.

In this work we limit our researches and discussion to the issue of breast cancer diagnosis based on fine needle biopsy (FNB) cytological images. The task at hand is to classify a case as benign or malignant. Many researchers have already

* This research was partially supported by National Science Centre in Poland.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_9

studied similar issues [13,15,11]. In our previous works we used a segmentation method based on the combination of adaptive thresholding, clustering in color space, fast marching and seeded watershed [10,8,5]. We tested the predictive accuracy on 450 images recorded by an analog video camera mounted atop of microscope. These experiments gives satisfactory results reaching 99.3% of correct classifications. Unfortunately, direct application of these methods for digital slides acquired with use of virtual microscopy system gave unsatisfactory predictive accuracy, mainly due to oversegmentation. To cope with this problem we applied classification driven multi-level thresholding to image segmentation [12]. Although regions extracted by this approach may not correspond to any medically interpretable objects but numerical experiments proved existence of implicit relationship between features of segmented regions and classes of breast cancer [9]. Here we propose a new system for exploring important regions of cytological images. The system is based on classification driven multi-level thresholding. In the first step, the HBMO algorithm is employed to find suboptimal thresholds in the sense of predictive accuracy. Then, classification tree is constructed using features generated by the suboptimal set of thresholds. Importance of image regions is determined based on classification tree structure. The system speed up the analysis of the cytological image by indicating which fragments of the image should be inspected in the first place. It can be also used to discover new, previously unknown features discriminating benign from malignant cases.

The remainder of this paper is organized as follows. In Section 2, material and methods are described. Section 3 gives the description of the experiments and study of the results obtained for implemented method. Concluding remarks are given in Section 4.

2 Material and Methods

2.1 Image Database

The database contains 50 slides of the cytological material obtained FNB. The material was collected from 50 patients of the clinic in Zielona Góra, Poland. The set contains 25 benign and 25 malignant lesions cases. Smears from the biological material were fixed in spray fixative and dyed with hematoxylin and eosin. Cytological preparations were then digitalized into virtual slides using the Olympus VS120 Virtual Microscopy System. The average size of the slides is approximately 200 000 $\times$ 100 000 pixels. On each slide a pathologist selected 11 areas which illustrated a sufficient amount of biological material. These areas were archived in RGB (8 bit/channel) TIFF files of size 1583 $\times$ 828pixels. The number of areas per one patient was recommended by the pathologists at the hospital and allows for a correct diagnosis. The image database contains 550 images (11 images per patient). Both malignant and benign sets contain the same number of images (275 images describing 25 cases). All cancers were histologically confirmed and all patients with benign disease were either biopsied or followed for a year.

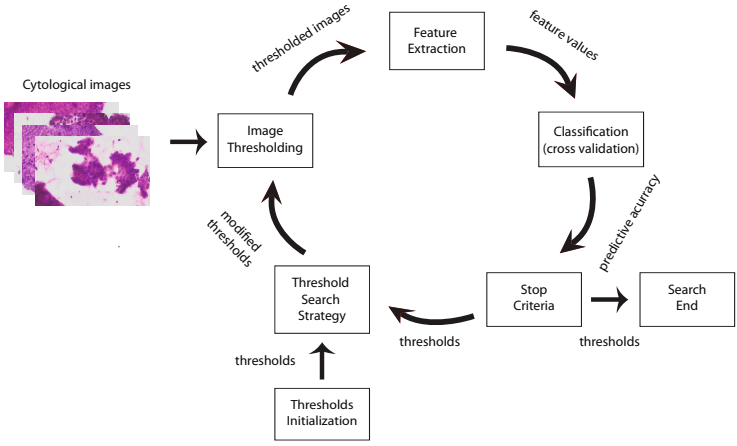


Fig. 1. Classification driven image thresholding

2.2 Multi-level Thresholding

Image thresholding boils down to discrete optimization problem with constraints. The key step of the algorithm is to determine set of thresholds corresponding to the maximum (or minimum) of the evaluation function:

$$T^* = \arg\max_T \{f(T) : T \in X \subset \mathbb{Z}^K\} \tag{1}$$

where $T = [t_1, t_1, \ldots, t_K]$ is a vector of thresholds, $f : \mathbb{Z}^K \rightarrow \mathbb{R}$ is an evaluation function, $X = \{T \in \mathbb{Z}^K : 0 \leq t_1 \leq t_2 \leq \ldots \leq t_K \leq L\}$ is a set of feasible solutions and L is a number of gray levels. In literature, a great variety of evaluation criteria have been already proposed [17,19]. Most of them are based on homogeneity measures. In cytological analysis, images are segmented to find the objects of interest and than their features are extracted to discriminate malignant from benign cases. Unfortunately, homogeneity based criterions can not guarantee obtaining high predictive accuracy because threshold selection is performed independently from the results of classification. To overcome this problem we propose to use a predictive accuracy to evaluate the thresholds. This approach originate from well known wrapper based feature selection strategy where predictive model is employed to score feature subsets [7]. In our approach classification tree is used to evaluate threshold sets [18,2]. The scheme of the approach is shown in Fig. 1. In the first step, thresholds are applied to segment images. Next, segmented regions are measured to compute features. Finally, images are classified and the predictive accuracy is estimated using n-fold cross-validation procedure [4]. Since the method trains a classification tree for each training subset, it is very computationally intensive. To deal with such computational effort we decided to use HBMO searching strategy. The HBMO algorithm belongs to the general class of swarm intelligence methods that models the behaviors of social insects. It is inspired by the marriage behavior of honey-bees [1,6]. The pseudo code of main steps of HBMO is given in Algorithm 4.

Algorithm 4. HBMO algorithm

```
Initialize randomly drones D and queen Q
while stop condition not met do
   while speed is above the threshold AND spermatheca is not full do
      Select randomly drone D_rand from set D
      Compute the probability of mating for drone D_rand
      if the drone D_rand passes the probabilistic condition then
         Add its sperm to the queens spermatheca
      end if
      Decrease the speed of queen
   end while
   for i = 1 : number of broods do
      Select randomly sperm P_rand from spermatheca
      Generate brood B_i based on P_rand and Q
   end for
   for j = 1 : number of mutations do
      Select randomly brood B_rand
      Apply mutation to B_rand
   end for
   Find the brood B_best with the highest value of objective function
   if objective function of B_best > objective function of queen Q then
      Replace current queen Q with B_best
   end if
end while
```

2.3 Discovering Important Regions of Image

There are many different approaches to the problem of discovering important regions of the image [3,14,16]. In our method, training images are segmented using suboptimal thresholds T^* determined by procedure described in Section 2.2. Then, features are computed for all images. Discriminative ability of individual features can be estimated by classification tree [2,18]. The full binary decision tree based on Gini diversity index as splitting criterion is generated. Then, the estimates of input feature importance are computed for tree by summing changes in the risk due to splits on every feature. At each node, the risk is estimated as node impurity and weighted by the node probability. Variable importance associated with this split is computed as the difference between the risk for the parent node and the total risk for the two children. Features with insignificant importance values are removed from the tree. Pruning is repeated until a significant deterioration in the predictive accuracy is observed. Structure of pruned tree is used to evaluate the region importance. Each node in the classification tree is associated with a single feature, and thus also with a single region in the image. The higher level the node occupies in the tree, the more important is region associated with him. The contours of important regions are mapped to original cytological image and in this way can be presented to pathologists. Sample results are presented in Fig. 5.

3 Numerical Experiments

To find optimal threshold set for 550 images described in Section 2.1, we performed the procedure described in Section 2.2. Six types of features are computed for each segmented region: area (A), perimeter (P), euler number (EN), variance in R channel (varR), variance in G channel (varG) and variance in B channel (varB). Thus each image is described by $n_f = 3(K + 1)$ features. For classifying images we use classification tree. Predictive accuracy is estimated using n-fold cross-validation procedure. There are 50 folds (the number of cases), and each fold consist of 11 images that belong to a single case. The images belonging to the same case are never at the same time in the training and testing set. The predictive accuracy is defined as the percentage ratio of successfully classified images to the total number of images. Remembering that HBMO belongs to class of stochastic global optimization algorithms, 10 runs of algorithm was performed for each number of thresholds. Computational effort was constant for each experiment and set to 500 evaluations. To find the optimal number of thresholds K^*, the searching procedure was performed for $K = 2, \ldots, 10$ thresholds. The

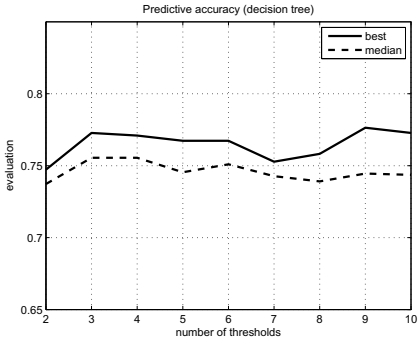


Fig. 2. The best and median classification results depending on the number of thresholds

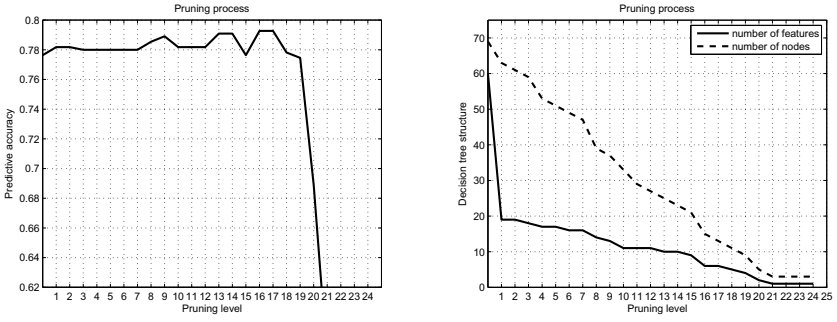


Fig. 3. Decision tree performance (left), and structure (right) over pruning process

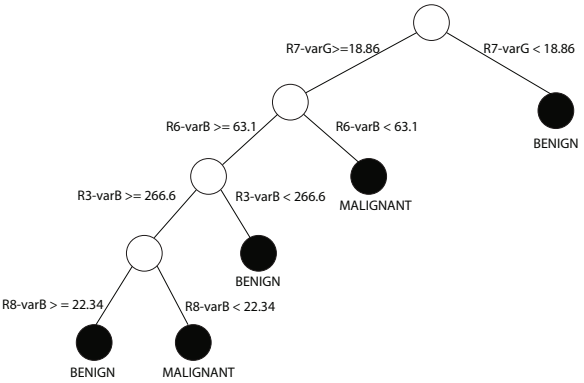


Fig. 4. The decision tree after pruning process (Rn-feature denotes the feature value for the $n-th$ region)

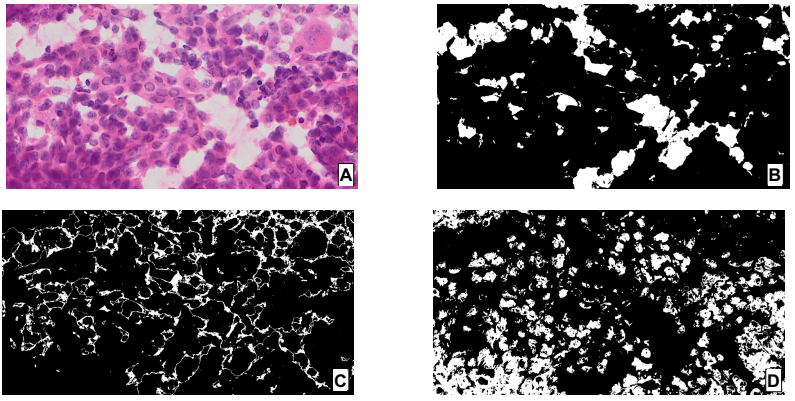


Fig. 5. Important regions: (A) original image, (B) binary mask for R7, (C) binary mask for R6, (D) binary mask for R3

relationship between the number of thresholds and predictive accuracy is shown in Fig. 2. The best classification accuracy 77.64% has been achieved by following thresholds $T = [57, 75, 113, 163, 174, 200, 225, 243, 249]$. Then, these thresholds are used to generate training data in order to induce the final classification tree. Next, tree is pruned to remove insignificant features. This process is described by Fig. 3. Pruned tree is shown in Fig. 4. The importance of image regions is determined based on position of nodes in the tree. Sample results are plotted in the form of binary masks presented in Fig. 5.

4 Conclusions

This study shows that proposed system of breast cancer diagnosis can discriminate benign from malignant cases with predictive accuracy equal to 77.64%.

Such result was achieved for image classifications accuracy. But single patient is described by 11 images and by applying majority voting for all of the classifier outcomes we can compute predictive accuracy for patient. This simple modification increases the predictive accuracy to 94%. However, it is too early to conclude that the system is ready to assist medical tasks by suggesting a diagnosis to pathologists. Higher predictive accuracy can be achieved if better quality feature space will be found. It seems that particularly important can be textural features of segmented objects. Further research will be performed to check this hypothesis. In its current form, the system can help less experienced pathologists by showing them image regions that are important from the point of view of breast cancer diagnosis. Our system can help also experienced pathologist to explore new unknown features discriminating benign from malignant cases. The main advantage of the proposed system is its generality because it can be relatively easily adapted to the analysis of microscopic images of any type. Although the training of the system is relatively computationally expensive but it is done off-line and only once. Image analysis is performed by tuned system very quickly.

References

1. Abbass, H.A.: MBO: Marriage in honey bees optimization - a haplometrosis polygynous swarming approach. In: Proceedings of the 2001 Congress on Evolutionary Computation, vol. 1, pp. 207–214. IEEE (2001)
2. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey (1984)
3. Choraś, R., Andrysiak, T., Choraś, M.: Integrated color, texture and shape information for content-based image retrieval, pattern analysis and applications. Pattern Analysis and Applications 10(4), 333–343 (2007)
4. Devijver, P.A., Kittler, J.: Pattern Recognition: A Statistical Approach. Prentice-Hall, London (1982)
5. Filipczuk, P., Krawczyk, B., Woźniak, M.: Classifier ensemble for an effective cytological image analysis. Pattern Recognition Letters 34(14), 1748–1757 (2013)
6. Horng, M.H.: A multilevel image thresholding using the honey bee mating optimization. Applied Mathematics and Computation 215(9), 3302–3310 (2010)
7. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artificial Intelligence 97(1), 273–324 (1997)
8. Kowal, M., Filipczuk, P.: Nuclei segmentation for computer-aided diagnosis of breast cancer. Int. J. Appl. Math and Comp. Sci. 24(1), 19–31 (2014)
9. Kowal, M., Filipczuk, P., Marciniak, A., Obuchowicz, A.: Swarm optimization and multi-level thresholding of cytological images for breast cancer diagnosis. In: Burduk, R., Jackowski, K., Kurzynski, M., Wozniak, M., Zolnierek, A. (eds.) CORES 2013. AISC, vol. 226, pp. 611–620. Springer, Heidelberg (2013)
10. Kowal, M., Filipczuk, P., Obuchowicz, A., Korbicz, J., Monczak, R.: Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. Computers in Biology and Medicine 43(10), 1563–1572 (2013)
11. Malek, J., Sebri, A., Mabrouk, S., Torki, K., Tourki, R.: Automated breast cancer diagnosis based on GVF-Snake segmentation, wavelet features extraction and fuzzy classification. Journal of Signal Processing Systems 55(1-3), 49–66 (2009)

12. Marciniak, A., Kowal, M., Filipczuk, P., Korbicz, J.: Swarm intelligence algorithms for multi-level image thresholding. In: Korbicz, J., Kowal, M. (eds.) Intelligent Systems in Technical and Medical Diagnostics. AISC, vol. 230, pp. 301–311. Springer, Heidelberg (2013)
13. Mazurek, P., Oszustowska-Mazurek, D.: From the slit-island method to the ising model: Analysis of irregular grayscale objects. Int. J. Appl. Math and Comp. Sci. 24(1), 49–63 (2014)
14. Mikolajczyk, K., Schmid, C.: Scale & affine invariant intrest point detectors. Int. J. Computer Vision 60(1), 63–86 (2004)
15. Niwas, S.I., Palanisamy, P., Sujathan, K., Bengtsson, E.: Analysis of nuclei textures of fine needle aspirated cytology images for breast cancer diagnosis using complex daubechies wavelets. Signal Processing 93(10), 2828–2837 (2013)
16. Ogiela, M.R., Trzupek, M., Tadeusiewicz, R.: Intelligent image content semantic description for cardiac 3d visualizations. Int. J. Engineering Applications of Artificial Intelligence 24(8), 1410–1418 (2011)
17. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Sys. Man. and Cyber. 9, 62–66 (1979)
18. Quinlan, J.R.: Induction of decision trees. Machine Learning 1(1), 81–106 (1986)
19. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. J. Electronic Imaging 13(1), 146–165 (2004)

Gaussian Mixture Model Based Non-Local Means Technique for Mixed Noise Suppression in Color Images

Maria Luszczkiewicz-Piatek

University of Lodz, Faculty of Mathematics and Computer Science,
Department of Applied Computer Science, Banacha 22, 90-238 Lodz, Poland
mluszczkiewicz@math.uni.lodz.pl

Abstract. In this paper a new approach to reduction of a mixture of Gaussian and random impulse noises in color images is presented. The proposed filtering scheme is based on the application of the Bilateral Filter in order to address the problem of impulse noise reduction by determining the rate of region homogeneity, for calculating the weights needed for the Non-Local Means (NLM) averaging operation. Gaussian Mixture Model approach is applied for determining similarity between local image regions. The proposed solution is capable to successfully suppress the mixed noise of various intensities, at a lower computational cost than NLM method, due to the adaptive choice of size of search window for similar local neighborhoods. Experimental results prove that the introduced design yields better results than the Non-Local Means and Anisotropic Diffusion techniques in the case of color images contaminated by strong mixed Gaussian and impulsive noise.

1 Introduction

Over the recent years, many spectacular technological achievements have revolutionized the way information is acquired and handled. Nowadays, more than ever, there is an exponentially growing number of color images being captured, stored and made available on the Internet and therefore the interest in color image enhancement is rapidly growing. Quite often, color images are corrupted by various types of noise introduced by malfunctioning sensors in the image formation pipeline, electronic instability of the image signal, faulty memory locations in hardware, aging of the storage material, transmission errors and electromagnetic interferences due to natural or man-made sources. The problem of the noise reduction is one of the most frequently performed image processing operation, as the enhancement of images or video streams corrupted by noise, is essential stage in order to facilitate further image processing steps.

Let us note, that there exist large collection of efficient methods designed to remove Gaussian noise. The significant leap was made by proposing Non-Local Means Filter (NLM) [1]. This idea is based on the estimation the original image using weighted mean along similar local patches. This filter is very efficient when

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_10

applied for restoring images corrupted by Gaussian noise, but fails in the presence of distortions introduced by impulsive noise. Similarly, the problem of impulse noise removal was also explored by many approaches e.g. [3,4,7].

Let us consider that although methods proved to be very effective in removing Gaussian noise generally fail when impulse noise is present. Moreover, methods successfully dealing with impulse noise usually lose their effectiveness when applied for other noise type removal.

In recent years a few filters were proposed for removing of a mixture of Gaussian and impulse noises [5,6,8], although such noises can take place quite often. One of the approaches to this problem is ROAD statistics [2] applied for detecting noisy pixels combined with Bilateral Filter [13] for Gaussian noise removal, resulting in Trilateral Filter (TriF) and its variations e.g. [7].

This work is focused on restoration of images corrupted by mixed Gaussian and impulse noise using the concept of the adaptive weighted averaging (i.e. NLM approach), where assigned weights help to determine the outlying observations, decreasing their influence on the filtering result based on the region homogeneity assessment, evaluated using Bilateral Filter approach.

The reminder of this paper is organized as follows. Section 2 introduces the concept of region homogeneity assessment using Bilateral Filter approach. In Section 3 the similarity between image patches using Gaussian Mixture Model is introduced. Non-Local Means filtering scheme is introduced in Section 4. Experimental setup and results are presented and discussed in Sections 5 and 6 accordingly.

2 Region Homogeneity

The proposed solution, based on NLM approach, averages image pixels on the basis of the similarity of their surrounding. However, if it would be the only criterion for evaluating new pixel value, only the influence of the Gaussian noise will be suppressed. Let us note, that even if the pixels, whose local neighborhood is similar to the local neighborhood of the pixel which is currently being processed are similar, the calculated average, can be significantly influenced if averaged pixels are corrupted by impulse noise. Therefore, the proposed approach overcomes this drawback by identifying those pixels which are outlying from their local surrounding. Thus, even if their local neighbourhood is similar to that of the processed pixel, they should be considered less important during the averaging process. This is achieved by the application of the *homogeneity maps* [10] based on Bilateral Filter approach. In details, each pixel is represented in the weight map, as the associated coefficient reflecting its similarity to the color of neighboring pixels with the relation to the spatial distance between them. In consequence, this approach assigns significantly larger weights to pixels belongings to large color regions, than to small ones, often reflecting unimportant details, artifacts or impulse noise. Thus, this approach provides information if analyzed pixel is significantly different than its neighbours, indicating that it can be corrupted by e.g. impulse noise or it is edge pixel.

In details, the weights assigned to the pixel at position (x, y) are computed according to the following scheme:

$$w_{x,y} = \frac{1}{n} \sum_{(i,j) \in W} \exp\left(-\frac{\| \boldsymbol{c}_{x,y} - \boldsymbol{c}_{i,j} \|}{h}\right)^{k_1} \cdot \exp\left(-\frac{d_{i,j}}{\delta}\right)^{k_2}, \quad (1)$$

where $\boldsymbol{c}_{i,j}$ and $\boldsymbol{c}_{x,y}$ denote the color pixels at positions (i, j) and (x, y) respectively, h is the color difference scaling parameter, $d_{i,j}$ is the Euclidean distance between the pixel at position (i, j) and (x, y), which is the center of the filtering window W and δ is a spatial normalizing parameter equal to the diameter of the square filtering window. The number of pixels n in W was set to be equal to 10% of the total number of pixels in the image and we assumed $k_1 = k_2 = 2$. For the color difference evaluation the $CIEDE2000$ color similarity measure [11] was used.

Fig. 1 presents the homogeneity maps, evaluated using the formula 1, for exemplary original images corrupted with mixture of Gaussian ($\sigma = 10$) and impulse ($p = 0.1$) noise. The lower values indicate that pixel neighbourhood is non-homogenous in terms of color.

Fig. 1. The color homogeneity maps (right in each pair) for original images corrupted with mixture of Gaussian ($\sigma = 10$) and impulse noise (p=0.1) (left in each pair)

3 Gaussian Mixture Modeling

Next step of the proposed methodology is to determine the similarity between pixel region on the basis of their Gaussian Mixture Models (GMM). Each local neighborhood is represented by GMM parameters. These parameters are compared in order to determine the region similarity.

The very important decision concerning the color image data modeling is the choice of the color space suitable for the retrieval experiments. In this paper the results were evaluated using the $CIE\ La^*b^*$ color space[12].

The first step in evaluating region similarity is to construct the histogram $H(x, y)$ in the $a-b$ chromaticity space defined as $H(x, y) = N^{-1}\sharp\{a_{i,j} = x, b_{i,j} = y\}$, 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 $\sharp$ denotes the number of samples in a bin and N is the number of color image pixels.

The next stage of the presented technique is the modeling of the color histogram using the Gaussian Mixture Models (GMM) and utilizing the

Expectation-Maximization (EM) algorithm for the model parameters estimation, [13].

Let us assume the following probabilistic model: $p(x|\Theta) = \sum_{m=1}^{M} \alpha_m p_m(x|\theta_m)$,
which is composed of M components and its parameters are defined as: $\Theta = (\alpha_1, \ldots \alpha_M, \theta_1, \ldots, \theta_M)$, with $\sum_{m=1}^{M} \alpha_m = 1$. Moreover, each p_m is a function of the probability density function which is parameterized by θ_m. Thus, the analyzed model consists of M components with M weighting coefficients α_m. Finally after derivations shown in [13] the model parameters are defined as:

$$\alpha_m^{k+1} = N^{-1} \sum_{i=1}^{N} p(m|x_i, \Theta^k), \quad \mu_m^{k+1} = \frac{\sum_{i=1}^{N} x_i \cdot p(m|x_i, \Theta^k)}{\sum_{i=1}^{N} p(m|x_i, \Theta^k)}, \tag{2}$$

$$\upsilon_m^{k+1} = \frac{\sum_{i=1}^{N} p(m|x_i, \Theta^k)(x_i - \mu_m^{k+1})(x_i - \mu_m^{k+1})^T}{\sum_{i=1}^{N} p(m|x_i, \Theta^k)} \quad , \tag{3}$$

where μ and υ denote the mean and variance, m is the index of the model component and k is the iteration number. The E (Expectation) and M (Maximization) steps are performed simultaneously, according to (2) and (3) and in each iteration, as the input data we use parameters obtained in the previous one, until convergence. The number of components are initially set to 20 and decreased in each iteration if associated weight α_m is less than 0.01.

4 Non-Local Means Algorithm

Let us assume the following image formation model $F(x, y) = I(x, y) + n(x, y)$, where x, y represents the pixel coordinates, I is the original image, F is the noise infested image, and $n(x, y)$ is the additive noise component respectively. The noise component is white, Gaussian noise distributed with zero mean and variance σ^2. The goal of the denoising is to obtain an estimate $\widehat{I}(x, y)$ from $I(x, y)$ using the observed image $F(x, y)$. The NLM algorithm computes the estimate of the original image according to the following equation: $\widehat{I}(x, y) = \frac{\sum_{k,l \epsilon I} W_{x,y}(k,l) F(k,l)}{\sum_{k,l \epsilon I} W_{k,l}}$. Finally, the evaluated filter output $\widehat{I}(x, y)$ is a weighted average of the image pixels. Let us underline that the NLM approach does not restrict the weighted average to only a local neighborhood of the processed pixel, but pixels of the entire image can be taken into account. The weights W are computed based on the differences of the two local regions centered at coordinates x, y and k,l as follows:

$$W_{x,y}(k, l) = \exp \frac{-\|F(\Omega_{x,y} - \Omega_{k,l})\|_a^2}{\varphi^2}, \tag{4}$$

where $\Omega_{x,y}$ and $\Omega_{k,l}$ are two windows centered at x, y and k, l respectively, $\| \cdot \|_a^2$ is the Gaussian weighted Euclidean distance, a is the standard deviation of the Gaussian kernel and φ is a parameter controlling the degree of filtering, chosen

experimentally, [1]. Let us note that the number of arithmetic operations needed is quite large. The solution proposed in this paper adaptively chooses the size of region S of searching of similar local neighborhoods $\Omega_{x,y}$ and $\Omega_{k,l}$ on the basis of the pixel nature (e.g. whether it is impulse noise corrupted or it is edge pixel).

5 Experimental Setup

In order to suppress noise using the proposed $GMM\text{-}NLM$ technique firstly the homogeneity map should be calculated using the scheme described in previous Section 2. Then the obtained values are analyzed. If the homogeneity map value $w(x, y)$ associated with the image pixel $F(x, y)$ is low it indicates that this pixel can be either influenced by impulse noise or is edge pixel. In both cases low value indicates that pixel is distinctively different that its surrounding. For noise suppression purposes it is important to differentiate between pixels which are impulse noise infested (need to be corrected) and belongs to edge (need to be preserved). Thus, the $GMM\text{-}NLM$ method examines the surrounding $\Upsilon_{x,y}$ of that pixel in homogeneity map in form of comparison between median value of the surrounding and the analyzed value. In case of these two values are distinctively different the search window $S_{x,y}$ in which the image patches $\Omega_{k,l}$ are compared to $\Omega_{x,y}$, is enlarged in order to find and take into account more similar local regions. On the contrary, if analyzed image pixel belongs to the edge, the enlargement of the search region $S_{x,y}$ will not improve the denoising results, because distant pixels, separated by the detected edge, possibly will not have similar neighborhoods to that of the analyzed pixel. Such approach is motivated by computation savings, i.e. less local regions to visit and compare.

The search window $S_{x,y}$ $(s_{x,y} \times s_{x,y})$ size is calculated on the basis of the formula $s_{x,y} = round(\xi \cdot r \cdot w(x, y) + \eta + 3)$ where ξ is the smaller of the width and length dimensions of the analyzed image, r is the scale ratio (here assumed $r = 5\%$), $w(x, y)$, $w\epsilon(0, 1)$ is the homogeneity coefficient evaluated using Eq.1, η is coefficient related to nature of pixel local non-homegeneity (i.e. $\eta = \kappa \cdot [w(x, y) - \tilde{\Upsilon}_{x,y}]$, for $\kappa = 10$). Then, the comparison of the local neighborhoods $\Omega_{k,l}$ for the image pixels in the search region $S_{x,y}$ to local neighborhood $\Omega_{x,y}$ is evaluated. The size of the image patch $\Omega_{x,y}$ $(\emptyset_{x,y} \times \emptyset_{x,y})$ is also related to homogeneity maps values $w_{x,y}$, such that $\emptyset_{x,y} = round(10 \cdot w_{x,y} + 3)$. Each local neighborhood Ω is is modeled using GMM technique described in Section 6 producing vector of model parameters compared by *Earth Mover's Distance* [14]. The measured EMD distance is defined as a minimum amount of work needed to transform one histogram into the other. As this method operates on signatures and their weights using GMM, 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 nature of the EMD measure, accepting slight shifts in color histogram rather than bin-by-bin comparison, is well suited in case of distortions introduced by Gaussian noise. Having the local patches similarities calculated, the new estimate of the pixel value $\widehat{I}(x, y)$ is evaluated on the basis of the formula 4 with region similarities calculated using approach introduced in this Section, where $\varphi = s$.

6 Noise Suppression Results

In order to test the proposed methodology the mixture of Gaussian and impulse noise were applied to the set of original images shown in Fig. 2. The efficiency of this approach was evaluated in terms of the visual quality of the restored image and also in terms of objective quality measures. Additionally, the proposed solution effectiveness was compared with the denoising NLM [1] and *Anisotropic* $Diffusion$ [15] methods.

Firstly, the proposed filtering scheme was tested on the standard test images: GIRL, LENA and PEPPERS (Fig. 2) corrupted with mixed Gaussian and impulse noise (salt & pepper in each channel). The test images were contaminated by: mixed Gaussian and impulsive noise of $\sigma = 10$ and $p = 0.1$, $\sigma = 20$ and $p = 0.2$, $\sigma = 30$ and $p = 0.3$, where p denotes the contamination probability. The restoration capabilities of the proposed solution in terms of visual quality for mixture noise of $\sigma = 0.1$ and $p = 0.1$ can be seen in Fig. 3. Let us note that significant noise suppression can be observed along with edge preserving. This Fig. also illustrates the ineffectiveness of the Non-Local Means technique when impulse noise is present.

(a) GIRL (b) LENA (c) PEPPERS

Fig. 2. Color test images

Fig. 4 illustrates the noise suppression capabilities of the proposed $GMM - NLM$ technique for mixture of Gaussian noise ($\sigma = 20$ and $\sigma = 30$) and impulse noise of various intensity $p = \{0.1, 0.2, 0.3\}$ in comparison with the original image $PEPPERS$.

The noise removal capabilities of the proposed solution was tested using the Peak Signal to Noise Ratio ($PSNR$) and the Mean Absolute Error (MAE). The Peak Signal to Noise Ratio is given as $PSNR = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$, where MSE (Mean Squared Error) is defined by $MSE = \frac{\sum_{i=1}^{N}\sum_{k=1}^{3}[O_k(x_i)-J_k(x_i)]^2}{3N}$, where k denotes $k-th$ color channel of analyzed pixel, O_k is original image, J denotes filtered version of original image and N is total number of image pixels. $PSNR$ measure is the impulse noise suppression of analyzed filtering method. The Mean Absolute Error (MAE) is used to estimate filter ability to preserve fine details and is calculated using the following formula: $MSE = \frac{\sum_{i=1}^{N}\sum_{k=1}^{3}[O_k(x_i)-J_k(x_i)]}{3N}$. The $PSNR$ and MAE values calculated for mixture of Gaussian ($\sigma = \{10, 20, 30\}$) and impulse noise ($p = \{0.1, 0.2, 0.3\}$) for test images shown in Fig. 2 are summarized in Fig. 5.

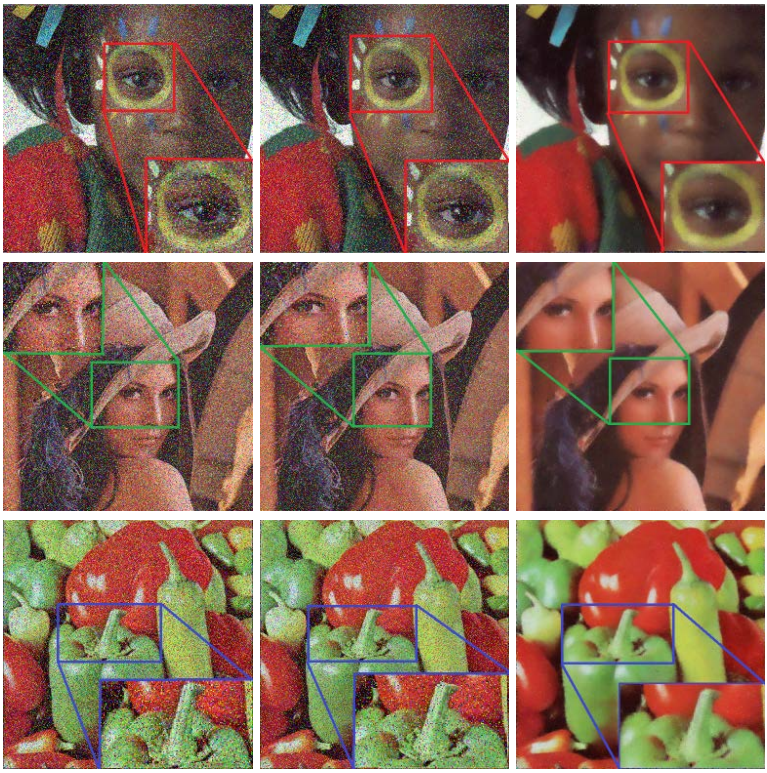


Fig. 3. The comparison between original images corrupted with mixture Gaussian ($\sigma = 20$) and impulse noise (p=0.2) (left) and its denoised versions usin Non-Local Means technique (middle) and proposed GMM-NLM method (right)

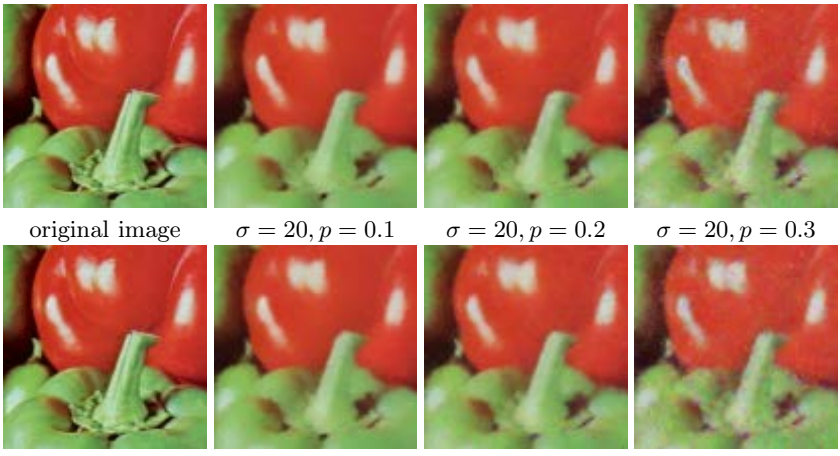


Fig. 4. The noise suppression results evaluated for mixture of Gaussian noise ($\sigma = 20$ in upper row and $\sigma = 30$ for bottom row) and impulse noise ($p = \{0.1, 0.2, 0.3\}$)

LENA							PEPPERS						
	PSNR			MAE				PSNR			MAE		
p−σ	10	20	30	10	20	30	p−σ	10	20	30	10	20	30
0.1	25.07	24.79	23.60	9.03	9.44	11.43	0.1	24.76	23.67	21.82	8.32	9.67	13.06
0.2	24.79	24.70	23.30	9.56	9.95	12.34	0.2	24.22	23.42	21.40	9.15	10.57	14.45
0.3	24.60	24.26	22.48	10.01	10.82	14.00	0.3	23.81	22.92	20.77	10.14	11.85	16.37

GIRL						
	PSNR			MAE		
p−σ	10	20	30	10	20	30
0.1	25.31	24.33	22.46	8.66	9.73	12.61
0.2	25.89	24.10	22.15	9.49	10.59	14.05
0.3	24.44	23.57	21.36	10.43	11.94	16.30

Fig. 5. The $PSNR$ and MAE values calculated for mixture of Gaussian ($\sigma = \{10, 20, 30\}$) and impulse noise ($p = \{0.1, 0.2, 0.3\}$) for test images shown in Fig. 2

LENA							PEPPERS						
	PSNR			MAE				PSNR			MAE		
(p, σ)	GMM	NLM	AD	GMM	NLM	AD		GMM	NLM	AD	GMM	NLM	AD
(0.1, 10)	25.074	19.51	25.24	9.03	9.22	9.16		24.76	18.54	24.33	8.32	10.49	10.69
(0.2, 20)	24.70	20.22	24.01	9.95	11.76	11.42		23.42	19.48	22.52	10.57	13.21	14.22
(0.3, 30)	22.48	21.09	22.67	14.08	13.93	14.43		20.717	20.84	20.83	16.37	16.59	18.06

Fig. 6. Comparison of $PSNR$ and MAE values evaluated for proposed solution, Non-Local Means [1] method and Anisotropic Diffusion [15]

The effectiveness of the proposed solution was also tested in comparison with Non-Local Means (NLM) method. In this approach the control parameters depend on the noise intensity and therefore they were experimentally chosen to provide the best possible noise suppression results. Proposed method efficiency was also compared with Anisotropic Diffusion (AD), [15]. The comparison results in terms of $PSNR$ and MAE were summarized in Fig. 6.

7 Conclusions

In this paper a new noise reduction method that combines the Non-Local Means averaging (for Gaussian noise removing) scheme with the region homogeneity assessment (for impulse noise suppression) is introduced. Our new noise reduction method, called GMM-NLM, outperforms significantly the NLM scheme in case of mixture noise, at a lower computational cost due to the adaptive choice of size of search window for similar local neighborhoods and the size of each image patch. Also, the noise parameters do not need to be estimated during the image processing. Moreover, this approach can decrease the computation cost in comparison with classical NLM approach. The proposed method was also compared with $Anisotropic\ Diffusion$ technique yielding better results.

Acknowledgment. This work was supported by the Polish National Science Center (NCN) under the Grant: DEC-2012/05/B/ST6/03428.

References

1. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: IEEE Conf. on Computer Vision and Pattern Recognition, USA, vol. 2, pp. 60–65 (2005)
2. Garnett, R., Huegerich, T., Chui, C., Wenjie, H.: A universal noise removal algorithm with an impulse detector. IEEE Transactions on Image Processing 14(11), 1747–1754 (2005)
3. Smolka, B.: Peer group switching fillter for impulse noise reduction in color images. Pattern Recognition Letters 31(6), 484–495 (2010)
4. Nikolova, M.: A variational approach to remove outliers and impulse noise. J. Math. Imag. Vis. 20(1), 99–120 (2004)
5. Tang, K., Astola, J., Neuvo, Y.: Nonlinear multivariate image filtering techniques. IEEE Transactions on Image Processing 4(6), 788–798 (1995)
6. Lukac, R., Smolka, B., Plataniotis, K.N., Venetsanopoulos, A.N.: Vector sigma lters for noise detection and removal in color images. Journal of Visual Communication and Image Representation 17(1), 1–26 (2006)
7. Dong, Y., Chan, R., Xu, S.: A detection statistic for random-valued impulse noise. IEEE Trans. Image Process. 16(4), 1112–1120 (2007)
8. Lamichhane, B.P.: Finite Element Techniques for Removing the Mixture of Gaussian and Impulsive Noise. IEEE Transactions on Signal Processing 57(7), 2538–2547 (2009)
9. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Sixth International Conference on Computer Vision, pp. 839–846 (1998)
10. Luszczkiewicz, M., Smolka, B.: Application of bilateral filtering and Gaussian mixture modeling for the retrieval of paintings. In: Proc. of ICIP, pp. 77–80 (2009)
11. Sharma, G., Wu, W., Dalal, E.N.: The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. Color Research and Application 30(1), 21–30 (2005)
12. Łuszczkiewicz-Piątek, M.: Which Color Space Should Be Chosen for Robust Color Image Retrieval Based on Mixture Modeling. In: S. Choras, R. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 55–64. Springer, Heidelberg (2014)
13. Bilmes, J.: A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. University of Berkeley, ICSI-TR-97-021 (1997)
14. Rubner, Y., Tomasi, C., Guibas, L.J.: The Earth Mover Distance as a Metric for Image Retrieval. International Journal of Computer Vision 40(2), 99–121 (2000)
15. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(7), 629–639 (1990)

Robust Image Retrieval Based on Mixture Modeling of Weighted Spatio-color Information

Maria Luszczkiewicz-Piatek[1] and Bogdan Smolka[2]

[1] University of Lodz, Faculty of Mathematics and Computer Science, Department of Applied Computer Science, Banacha 22, 90-238 Lodz, Poland
mluszczkiewicz@math.uni.lodz.pl
[2] Silesian University of Technology, Department of Automatic Control, Akademicka 16 Str., 44-100 Gliwice, Poland
bogdan.smolka@polsl.pl

Abstract. In this paper we propose a novel approach to color image retrieval. Color information is modeled using Gaussian mixtures and incorporates the information on the spatial distribution of the color image pixels utilizing the Dijkstra algorithm. The proposed algorithm has high indexing performance and operates on model of low dimensionality. Thus, the proposed method needs only the adjustment of Gaussian Mixture Model parameters for efficient color image retrieval. The proposed method is extensively tested on Corel and Wang dataset. The results demonstrate that proposed framework is more efficient than other methods when images are subjected to lossy coding such as JPEG method.

1 Introduction

In the recent years, the amount of digital images available has grown rapidly due to affordable digital cameras and high-speed Internet connections. Those factors have created a convenient way to generate and publish everyday enormous amount of visual content available worldwide by a growing number of users. Thus, there is a huge demand for image management tools designed for purposes of storing, browsing and searching in large multimedia databases.

In Content-Based Image Retrieval (CBIR) systems, the proper image description is a very important element assessing the similarities among images. Thus, image descriptors can be classified depending on the image property such as color or texture. In CBIR systems, the searching process is based on user given queries, which is often an image, represented by its properties encoded into features vectors, based on concepts of the i.e. color histogram, the color coherence vectors, the color co-occurrence matrix, vector quantization, and then compared in order to find similar images. Also the spatial organization of colors has been explored in form of spatial statistics between color pixels, such correlograms or some filter responses [1,2,3].

Let us note that due to problems related to the high dimensionality of data [5], it is preferable to employ fewer features to accurately represent image information and reduce computational cost, without deteriorating discriminative capability. This approach attracted great attention in recent years, [6].

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_11

This paper addresses the problem of similar image retrieval, when query image is given, on the basis of the image signature composed of Gaussian Mixture Model parameters. Although several proposed image retrieval techniques utilize the *Gaussian Mixture Model* (GMM) as color distribution descriptor [7,8,9], the aspect of the distortions caused by the lossy compression was not taken into account. The spatial arrangement of the color regions present in the analyzed image is incorporated during the color information modelling in form of color homogeneity information evaluated using Dijkstra algorithm, [10]. Due to the smoothing properties of the Gaussian mixtures, this approach is robust to distortions introduced by lossy codding, i.e. $JPEG$ compression scheme. The paper is organized as follows. The details of the proposed technique are described in Section 2, it also focuses on the evaluation of spatial arrangement of the color pixels within the scene depicted in the image using the Dijkstra algorithm. Next Section 3 presents the experimental setup of the retrieval scheme based on the proposed solution. The comparison of the experimental results for various compression ratios is also presented in Section 4. Finally, Section 5 concludes the concepts and results presented in this paper.

2 Gaussian Mixture Modeling

The first and very important decision concerning the color image data modeling using any technique is the choice of the color space suitable for the retrieval experiments [11]. In this paper we are using the $CIE\ La^*b^*$ color space. 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 the component equal to x and b component equal to y, the symbol # denotes the number of samples in a bin and N is the number of color image pixels. The next stage of the presented technique is the modeling of the color histogram using the *Gaussian Mixture Model* (GMM) and utilizing the Expectation-Maximization (EM) algorithm for the model parameters estimation as described in details in [12], using following formulae:

$$\alpha_m^{k+1} = N^{-1} \sum\nolimits_{i=1}^{N} p(m|x_i, \Theta^k), \quad \mu_m^{k+1} = \frac{\sum_{i=1}^{N} x_i \cdot p(m|x_i, \Theta^k)}{\sum_{i=1}^{N} p(m|x_i, \Theta^k)}, \tag{1}$$

$$\upsilon_m^{k+1} = \frac{\sum_{i=1}^{N} p(m|x_i, \Theta^k)(x_i - \mu_m^{k+1})(x_i - \mu_m^{k+1})^T}{\sum_{i=1}^{N} p(m|x_i, \Theta^k)} \quad , \tag{2}$$

where μ and υ denote the mean and variance, m is the index of the model component and k is the iteration number. The E (Expectation) and M (Maximization) steps are performed simultaneously, according to (1) and (2) and in each iteration, and the input data we use parameters obtained in the previous one. The main idea of the application of the GMM technique lies in the highly desirable

properties of this approach. The inherent feature of the GMM enables to approximate the distorted color histogram of the color image subjected to lossy compression, which is obtained through limited model complexity (7 components) and number of iterations (75) of E-M algorithm, as shown in [15,14,13,16]. The lossy compression causes a loss of color information resulting in discontinuities in the chromaticity histogram. The proposed methodology counteracts this effect by smoothing the histogram in order to reconstruct its original shape, which is a basis for the effective image retrieval, which lack other approaches based on mixture modeling [17].

Thus, in order to reflect the differences in image content and also to improve the retrieval efficiency of the proposed retrieval scheme, the concept of using *Dijkstra algorithm* [10], was incorporated. Its use provides a possibility to emphasize in the $a-b$ color histogram all the pixels belonging to homogenous color areas. In general, its construction reflects topological *distance* and color *similarity* of image pixels, evaluated using $CIEDE2000$ color similarity measure [18], represented by *optimal paths* between central pixel and each of its neighbors in processing window.

Using the concept of *Dijkstra algorithm*, each pixel of the original image is taken into the $a-b$ histogram with the weight evaluated on the basis of the sum of the costs of the "cheapest" paths to pixels belonging to processing window W, where cost of visiting neighboring pixels is calculated as difference of their colors using $CIEDE2000$ (Fig. 1). Thus, chromaticity histogram is in fact a histogram of weights associated to color pixels. Each bin value is the sum of the weights of the pixels of the particular $a-b$ values. The idea of evaluating color region homogeneity is based on treating processing window W (assigned as $h_1 = 10\%$ and $h_2 = 10\%$ of image height and width respectively) centered at pixel $y(i,j)$ and consisted of K pixels, as a directed graph.

The Dijkstra algorithm creates the path of the lowest total cost from the center pixel $y(i,j)$ of the processing window W to the pixel $y(i,j)^{(k)} \epsilon W$ and assigns the optimal cost $p(i,j)^{(k)}$ The weight associated with pixel $y(i,j)$, is based on the average of all optimal path cost evaluated for pixels of processing window W: $w(i,j) = \exp\left(-\frac{\sum_{k=1}^{K} p(y(i,j)^{(k)})^2}{h_1 \cdot h_2}\right)$. Thus, $w(i,j)$ can be treated as similarity measure between central pixel $y(i,j)$ and it surrounding $y(i,j)^{(k)} \epsilon W$. Let us note that, the more homogenous the pixel neighborhood, the lower is the the sum of costs of the optimal paths to all window pixels, and thus the larger is the weight value $w(i,j)$. In case of regions presenting significant color changes the total cost of optimal paths will be larger due to observed color differences among pixels.

3 Experimental Setup

In order to test the proposed methodology we subjected to lossy coding ($JPEG$) the database of Wang [4] consisting of 1000 natural images, of 10 semantic categories, and Corel dataset of 10 000 color images. For the purposes of the histogram comparisons, several distance or similarity measures were used (denoted

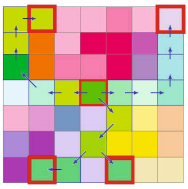


Fig. 1. Exmplary optimum paths with minimum costs for Dijkstra algorithm using $CIEDE2000$ as a color similarity measure

as d) and computed between the histogram of the evaluated image (denoted as H) and the 2D surface generated by the GMM model of its histogram, (denoted as C). Among them are geometric measures treating objects as vectors in a multi-dimensional space and computing the distance between two data points on the basis of pairwise comparisons in each dimension (Minkowski family distances such as L_1 and L_2 and Canberra metric). Information-theoretic measures (such as Jeffrey divergence) derived from Shannon's entropy theory and statistical measures (such as χ^2) compare objets not in pairwise manner but rather as a distributions. Following formulas describe applied similarity measures:

- L_1 (D_{L_1}) : $d(H,C) = \sum_{i=1}^{\eta}\sum_{j=1}^{\eta}|H_{i,j} - C_{i,j}|$,
- L_2 (D_{L_2}): $d(H,C) = \left(\sum_{i=1}^{\eta}\sum_{j=1}^{\eta}|H_{i,j} - C_{i,j}|^2\right)^{\frac{1}{2}}$,
- Canberra: (D_C) $d(H,C) = \left(\sum_{i=1}^{\eta}\sum_{j=1}^{\eta}\frac{|H_{i,j}-C_{i,j}|}{|H_{i,j}|+|C_{i,j}|}\right)$,
- Jeffrey Divergence (D_J): $d(H,C) = \left(\sum_{i=1}^{\eta}\sum_{j=1}^{\eta}H_{i,j}\cdot\log\frac{H_{i,j}}{\eta_{i,j}} + C_{i,j}\cdot\log\frac{C_{i,j}}{\eta_{i,j}}\right)$,
- χ^2 (D_{χ^2}): $d(H,C) = \left(\sum_{i=1}^{\eta}\sum_{j=1}^{\eta}\frac{(H_{i,j}-\eta_{i,j})^2}{\eta_{i,j}}\right)$,

where $\eta_{i,j} = \frac{H_{i,j}+C_{i,j}}{2}$. For the evaluation of the difference between two histograms we also used the the *Earth Mover's Distance* (EMD) similarity measure, [19]. 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 GMM, we assigned as signature values the *mean* of each component and for the *signature weight* the weighting coefficient of each Gaussian in the model.

4 Evaluation of the Method Efficiency

In order to present retrieval observations, the *Precision* and *Recall* measures are employed. In more details, *Precision* is the fraction of retrieved images that are relevant, while Recall is the fraction of relevant instances that are retrieved. Fig. 2 illustrates the $Precision - Recall$ plots evaluated for the entire analyzed

Wang database, i.e. each of the database image was used as a query. In these experiments the criterion of the successful retrieval was the membership to the same thematic group as the query, not necessarily sharing the same color palette. Therefore, the main aim of the presented plots and numerical results is to show that proposed technique can produce meaningful results regardlessly to the rate of information loss associated with various compression methods. Fig. 2 compares the retrieval efficiency for several similarity measures. It can be noticed that there is no significant difference between retrieval accuracy for various compression ratios. The evaluated *Precision* values for each of 1000 queries were averaged (for each corresponding *Recall* value) producing the results shown in Fig. 3. The use of EMD similarity measure produced the best results, and moreover it offers compact color description contrary to other measures operating on histograms.

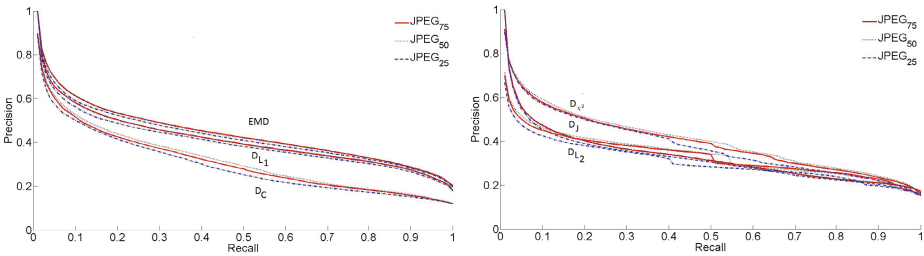


Fig. 2. The $Precision - Recall$ plots evaluated for Wang database of 1000 natural color images subjected to lossy coding using compression rates: 75% ($JPEG_{75}$), 50% ($JPEG_{50}$) and 25% ($JPEG_{25}$) for various similarity measures: D_{L_1}, D_{L_2}, D_C metrics, D_J, D_{χ^2} and EMD

Dataset	EMD	D_C	D_{L_1}	D_{L_2}	D_J	D_{χ^2}
$JPEG_{75}$	0.4125	0.3099	0.4384	0.3302	0.3324	0.3941
$JPEG_{50}$	0.4109	0.3194	0.4370	0.3304	0.3353	0.4030
$JPEG_{25}$	0.4008	0.2948	0.4272	0.3214	0.3062	0.3791

Fig. 3. The average *Precision* values evaluated for Wang database of 1000 natural color images subjected to $JPEG$ lossy coding using compression rates: 75% ($JPEG_{75}$), 50% ($JPEG_{50}$) and 25% ($JPEG_{25}$) for various similarity measures: D_{L_1}, D_{L_2}, D_C metrics, D_J, D_{χ^2} and (EMD)

Let us note, that *Precision* and *Recall* tend to ignore the ranking of retrieved images, e.g. *Precision* of 0.5 indicates that half of the retrieved images is relevant to the given query but without any further information if relevant images are the first half of the retrieved set or are the second half. In order to counteract this, it is advised to measure *Precision* and *Recall* at the specified points e.g. at the answer set of 1 ($\hat{P}_1$), 5 ($\hat{P}_5$), 10 ($\hat{P}_{10}$), and 25 ($\hat{P}_{25}$) images. These measurements (Fig. 4) are summarized to give the better view of the behavior of the analyzed

	EMD				D_C				L_1 (D_{L_1})			
Dataset	$\hat{P}_1$	$\hat{P}_5$	$\hat{P}_{10}$	$\hat{P}_{25}$	$\hat{P}_1$	$\hat{P}_5$	$\hat{P}_{10}$	$\hat{P}_{25}$	$\hat{P}_1$	$\hat{P}_5$	$\hat{P}_{10}$	$\hat{P}_{25}$
$JPEG_{75}$	1	0.6569	0.5810	0.4794	1	0.6569	0.5810	0.4794	1	0.6569	0.5810	0.4794
$JPEG_{50}$	1	0.6531	0.5795	0.4780	1	0.6531	0.5795	0.4780	1	0.6531	0.5795	0.4780
$JPEG_{25}$	1	0.6430	0.5628	0.4647	1	0.6430	0.5628	0.4647	1	0.6430	0.5628	0.4647

	D_{L_2}				D_J				D_{χ^2}			
Dataset	$\hat{P}_1$	$\hat{P}_5$	$\hat{P}_{10}$	$\hat{P}_{25}$	$\hat{P}_1$	$\hat{P}_5$	$\hat{P}_{10}$	$\hat{P}_{25}$	$\hat{P}_1$	$\hat{P}_5$	$\hat{P}_{10}$	$\hat{P}_{25}$
$JPEG_{75}$	1	0.5574	0.4733	0.3814	1	0.5576	0.4716	0.3800	1	0.5482	0.4541	0.3714
$JPEG_{50}$	0.6996	0.5003	0.4496	0.3926	0.7140	0.5161	0.4626	0.4011	0.663	0.4734	0.4246	0.3668
$JPEG_{25}$	0.9097	0.6552	0.5770	0.4781	0.9116	0.6653	0.5919	0.4872	0.8956	0.6480	0.5706	0.4743

Fig. 4. The average *Precision* values evaluated for Wang database of 1000 natural color images subjected to lossy coding using compression rates: 75% ($JPEG_{75}$), 50% ($JPEG_{50}$) and 25% ($JPEG_{25}$) for various similarity measures: D_{L_1}, D_{L_2}, D_C metrics, D_J, D_{χ^2} and EMD at points corresponding to 1, 5, 10 and 25 retrieved images

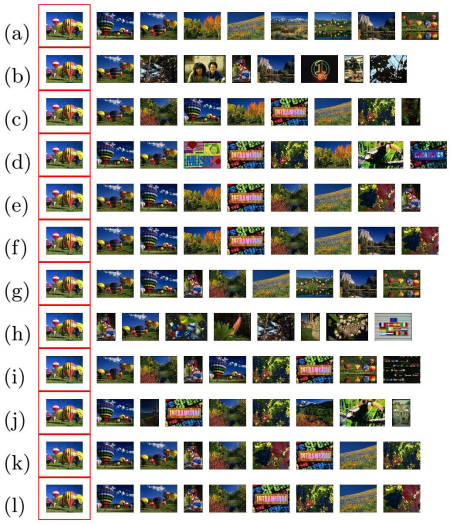


Fig. 5. The comparison of the retrieval results performed for Corel 10k database obtained for various similarity measures: EMD (a,g), *Canberra* (b,h), L_1 (c,i), L_2 (d,j), *Jeffrey divergence* (e,k), χ^2 (f,l). The results were evaluated for images transformed to GIF (a-f) and JPG_{25} (g-l)

retrieval scheme for applied color spaces. The EMD similarity measure produced the best results.

Fig. 4 illustrates the comparison between retrieval results evaluated for exemplary image of the Corel database of 10 000 color images for the various similarity measures for lossy coded versions of original images (25% original file size) and GIF (using dithering) method. It can be noticed that the evaluated results preserve the spatio-chromatic structure of the query image. The L_2 similarity measure yield the worst results, and should not be used for image indexing purposes.

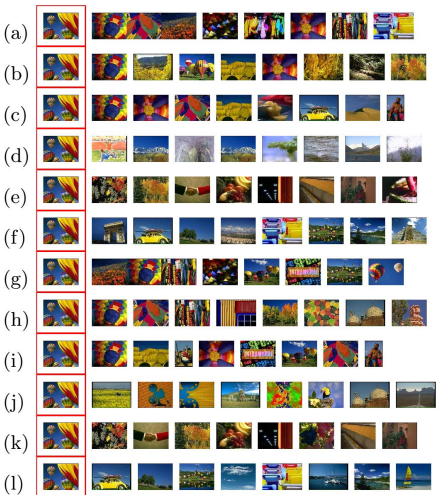


Fig. 6. The comparison of the retrieval results performed for Corel 10k database obtained for various retrieval methods: GMM with Dijkstra weighting applied for the images represented in the $CIE\ La^*b^*$ color space (a,g), CEDD [20] (b,h), FCTH [20] (c,i), MPEG-7 SCD (d,j), MPEG-7 CLD (e,k), Autocorrelogram (f,l). The results were evaluated for images transformed to GIF (a-f) and JPG_{25} (g-l)

The retrieval results evaluated for EMD similarity measure, were compared to those obtained using the well known, image retrieval application img(Rummager) [20]. This system contains several methods, from which the following were chosen: 60 bin histogram (CEDD), 192 bin histogram FCTH, MPEG-7 SCD and MPEG-7 CLD, and Autocorrelogram. Fig. 4 presents the retreval results for exemplary image of the Corel database transformed using GIF and JPG_{25} schemes.

When analyzing the performance of image retrieval techniques in comparison to scheme proposed in this paper, (Fig. 4) it can be noticed that, the color palette query image is not preserved (especially when GIF method was applied), the spatial arrangement the retrieved images is not always taken into account.

5 Conclusion

Many retrieval system rely on the color composition of the analyzed images. Although, this approach seems to be generally correct and effective, one must be aware of the problem of accurately managing the vast amount of visual information. The methods operating on chromaticity histograms could be severely disabled as color palettes of lossy compressed images can differ, providing misleading conclusions. As shown in this paper, such a comparison produces incorrect results when the retrieval process is evaluated not on the original images but on their compressed versions. Thus, there is an urgent need for the evaluation of techniques overcoming that undesirable phenomenon. Such a method is described in this paper.

The main contribution of this work is the adaptation of the Gaussian Mixture Models and the application of the Dijkstra algorithm for the purposes of the distorted chromaticity histogram approximation. The proposed scheme enables the retrieval system not only to take into account the overall image color palette but also to consider the the color dispersion understood as spatial arrangement of image colors. Let us note, that due to the fact that proposed solution preserves regions of homogenous color it can be applied also for retrieval of image of similar background content.

The satisfactory results were achieved independently on the applied compression scheme. Therefore, the presented results proved the hypothesis that the loss of color information caused by lossy coding can be efficiently counteracted, preserving spatial arrangement of colors in an image, providing successful retrieval results.

Acknowledgment. This work has been supported by The National Science Centre under SONATA grant no. UMO-2011/01/D/ST6/04554.

References

1. Ciocca, G., Schettini, L., Cinque, L.: Image Indexing and Retrieval Using Spatial Chromatic Histograms and Signatures. In: Proc. of CGIV, pp. 255–258 (2002)
2. Lambert, P., Harvey, N., Grecu, H.: Image Retrieval Using Spatial Chromatic Histograms. In: Proc. of CGIV, pp. 343–347 (2004)
3. Hartut, T., Gousseau, Y., Schmitt, F.: Adaptive Image Retrieval Based on the Spatial Organization of Colors. Computer Vision and Image Understanding 112, 101–113 (2008)
4. Wang, J.Z., Li, J., Wiederhold, G.: SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries. IEEE Trans. Patt. Anal. Mach. Intel. 9, 947–963 (2001)
5. Fodor, I.K.: A survey of dimension reduction techniques. Technical Report UCRL- ID-148494, Lawrence Livermore National Laboratory, US Department of Energy (2002)
6. Wu, P., Manjunath, B.S., Shin, H.D.: Dimensionality Reduction for Image Retrieval. In: Proc. of ICIP (2000)
7. Perronnin, F., Liu, Y., Sanchez, J., Poirier, H.: A similarity measure between unordered vector sets with application to image categorization. In: Proc. of CVPR, pp. 3384–3391 (2008)
8. Xing, X., Zhang, Y., Gong, B.: Mixture Model Based Contextual Image Retrieval. In: Proc. of CIVR 2010, pp. 251–258 (2010)
9. Beecks, C., Ivanescu, A.M., Kirchhoff, S., Seidl, T.: Modeling Image Similarity by Gaussian Mixture Models and the Signature Quadratic Form Distance. In: Proc. of ICCV, pp. 1754–1761 (2011)
10. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numerische Mathematik 1, 269–271 (1959)
11. Łuszczkiewicz-Piątek, M.: Which Color Space Should Be Chosen for Robust Color Image Retrieval Based on Mixture Modeling. In: Choras, R.S. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 55–64. Springer, Heidelberg (2014)

12. Bilmes, J.: A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. University of Berkeley, ICSI-TR-97-021 (1997)
13. Luszczkiewicz-Piatek, M., Smolka, B.: Selective color image retrieval based on the Gaussian mixture model. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P., Zemčík, P. (eds.) ACIVS 2012. LNCS, vol. 7517, pp. 431–443. Springer, Heidelberg (2012)
14. Luszczkiewicz, M., Smolka, B.: Application of bilateral filtering and Gaussian mixture modeling for the retrieval of paintings. In: Proc. of ICIP, pp. 77–80 (2009)
15. Luszczkiewicz-Piatek, M., Smolka, B.: Effective color image retrieval based on the Gaussian mixture model. In: Schettini, R., Tominaga, S., Trémeau, A. (eds.) CCIW 2011. LNCS, vol. 6626, pp. 199–213. Springer, Heidelberg (2011)
16. 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)
17. Avrithis, Y., Kalantidis, Y.: Approximate Gaussian Mixtures for Large Scale Vocabularies. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 15–28. Springer, Heidelberg (2012)
18. Sharma, G., Wu, W., Dalal, E.N.: The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. Color Research and Application 30(1), 21–30 (2005)
19. Rubner, Y., Tomasi, C., Guibas, L.J.: The Earth Mover Distance as a Metric for Image Retrieval. International Journal of Computer Vision 40(2), 99–121 (2000)
20. Chatzichristofis, S.A., Boutalis, Y.S., Lux, M.: Img (rummager): An interactive content based image retrieval system. In: Proc. of the 2nd International Workshop on Similarity Search and Applications (SISAP), pp. 151–153. IEEE Computer Society (2009)

Automatic Classification of Tahitian Pearls

Martin Loesdau, Sébastien Chabrier, and Alban Gabillon

Laboratoire d'Excellence CORAIL Géopôle du Pacifique Sud EA4238 Université de la Polynésie française, Tahiti
{martin.loesdau,sebastien.chabrier,alban.gabillon}@upf.pf

Abstract. Tahitian pearls are currently classified manually by pearl experts. This process is time consuming and furthermore subjective for some of the quality parameters such as colour, form or lustre. In this paper we present our ongoing work to implement an automatic classification of Tahitian pearls out of images. For this purpose different image segmentation and machine learning methods are used. In the following sections we explain our methodology, show first results for several sub-steps and give new ideas for greylevel edge detection and colour classification.

1 Introduction

Grown in the clear warm water of the Polynesian lagoons, the Tahitian pearl has fascinated humans since centuries. On the international market the pearl is known by the name 'Queen of Pearls', due to its pureness, its colour diversity and its brilliant lustre. As a result of its high reputation, the pearl is an important source of income for the French Polynesian government. During the years 2008 to 2010 the pearl contributed averagely with over 60% to the total exports of French Polynesia with an annual average value of over 90 million Euros. To keep the high reputation of the Tahitian pearl the local government implemented a quality grading sy stem, not allowing exporting pearls which does not satisfy the quality standards. To define the quality of a pearl several parameters are evaluated (form, size, defects, lustre, colour and minimal nacre thickness). This evaluation is nowadays done manually by pearl experts from the 'Direction de Ressources Marines et Minieres (DRMM)' [1]. This manual classification is problematic in two ways: First of all, the pure amount of over 1 million exported pearls per year is difficult to handle, as each pearl has to be classified separately. Second, the classification of several parameters, such as the form, the colour and the lustre is highly subjective. The goal of our project is to automatically classify pearls out of images, using image pr ocessing and machine learning techniques. An automatic classification gives the possibility to accelerate the quality evaluation process as well as to minimize the influence of subjectivity during the human classification. A third reason to implement an automatic classification is to support the scientific work of the 'Centre IFREMER du Pacifique' [2]. One of their research groups is analysing the factors that influence the quality of a pearl. Therefor correlation calculations between growing parameters (geographic location of the pearl farm, species of the breeding oyster, etc.) and the pearls

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_12

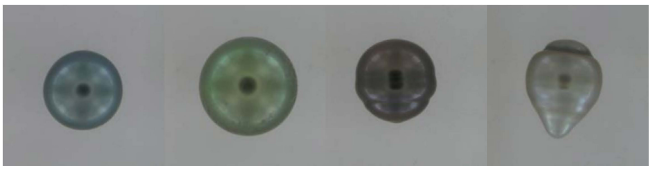


Fig. 1. Colour images of four Tahitian pearls

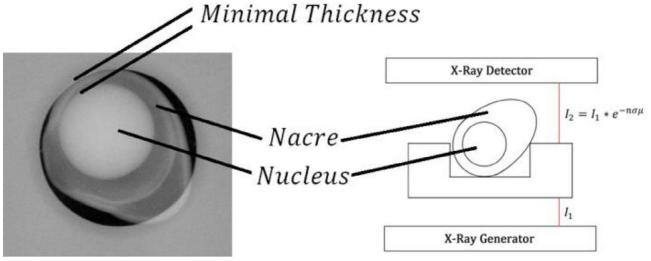


Fig. 2. X-Ray image of a Tahitian pearl (left) and a schema of the capturing (right)

quality are done. As a correlation calculation depending on subjectively classified quality parameters is not very trustworthy, an automatic classification of the parameters is preferable.

To analyse the quality parameters of a pearl, we use two different kinds of images. To determine the 'outer' parameters colour, size, form, lustre, and defects, colour images of the surface of a pearl were taken (Fig. 1). For each pearl four images were shot, to capture as much surface information as possible.

The minimal thickness of the nacre on the other hand cannot be evaluated from the outside. For this purpose X-Ray images are taken at the DRMM. In Fig. 2 on the left an X-Ray image of a Tahitian pearl can be seen.

The light round object in the middle of the pearl is the pearls nucleus. It is an artificially formed sphere which is inserted in a pearl oyster (Pinctada margaritifera in French Polynesia). Around this nucleus the oyster deposits a form of calcium carbonate, the nacre. The minimal thickness of the nacre is accordingly the minimal distance between the outer surface of the nucleus and the outer surface of the pearl. Currently this distance is measured manually by an employee of the DRMM (a video explaining this process in detail can be found under [3]).

In the following two sections we present our methodology to automatically determine and measure the named quality parameters. First results and general ideas regarding greylevel and colour image segmentation are given as well. In the fourth chapter our approach for colour classification is shown, followed by the conclusion in the fifth section.

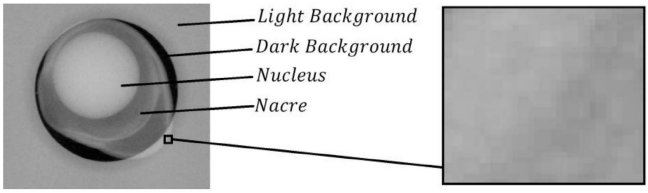


Fig. 3. Four different regions in the X-Ray image have to be segmented to measure the minimal nacre thickness (left). On the right, a detail showing the transition between the pearls nacre and the light background can be seen

2 2 X-Ray Images

To determine the minimal thickness of the nacre, the minimal distance between the artificial nucleus within the pearl and the outer surface of its surrounding nacre has to be measured. For this purpose X-Ray images of the pearl are taken. The resulting images contain 4 major regions: The light background, the dark background, the nucleus and the nacre (Fig. 3).

To measure the distance between the nucleus and the nacre the 4 regions have to be segmented. Over the recent years many approaches for greylevel image segmentation have been published. The classical approaches include techniques like Canny, Prewitt or Sobel edge detection [4,5,6,7,8]. Although these techniques are a standard in greylevel edge detection, some major drawbacks are known. These classical operators look pixel-wise for sharp intensity gradients in an image. A transition between different objects in the image, which ranges over many pixels without any sharp intensity difference between single pixels, is hard to detect (as for example between the nacre over the light background and the light background itself in Fig. 3 on the right). A threshold to separate edges from objects has to be defined for the whole image, which doesn't take into account a heterogenic cha racter of an image with objects of different homogeneity and edges of different i ntensity. The obtained edges are not necessarily connected, meaning a post pr ocessing of the obtained edges is necessary. Our current work evaluates the possibilities of a new theory for greylevel image edge detection. Our approach is to aggregate pixel of a monotonous gradient together and afterwards analyse these gradients. The advantage over the classical techniques is that transitions between objects without a sharp pixel-wise gradient might as well be detected. In Fig. 4 on the left a detail of intensity values of a column of pixels of the X-Ray image can be seen. The rectangles mark the pixels that belong to a monotonous ascending or descending gradient.

On the right side of the image is a binary result of thresholding every obtained gradient of the original X-Ray image column- and row-wise. The threshold was set to a minimal gradient value of 15. It can be seen that transitions between regions that do not include a sharp pixel-wise gradient can be detected as well. Anyhow a global threshold for an image is to the authors' opinion generally not a satisfying approach, as it doesn't respect a heterogenic character of an image.

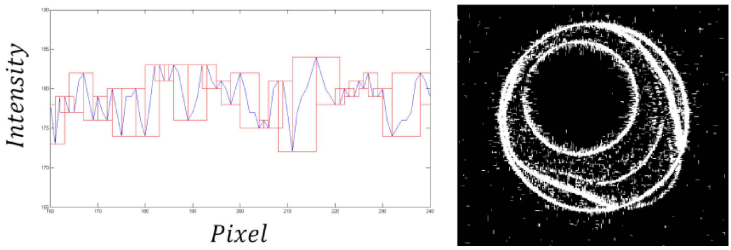


Fig. 4. Grouped pixels that form part of a monotonous gradient (left). On the right a binary image that shows row- and column-wise gradients bigger than 15

Fig. 5. Different segmented regions of a colour image of a Tahitian pearl and the quality parameters that are described in these regions

Our current research deals with finding solutions to discriminate between objects and edges independently of the gradients' size.

3 Colour Images

As mentioned, the colour images of the pearls are used to determine the parameters size, form, lustre, defects and colour. For this purpose multiple segmentations have to be done. In the case of the colour images, the crucial regions can be detected with less effort compared to the X-Ray images. Multiple global thresholds of the Intensity value obtained by converting the image to the HSI space lead to the results shown in Fig. 5. On the right side of each segmented area is the pearls quality parameter this region defines.

For each parameter feature vectors, mathematically describing a quality parameter, have to be calculated. These vectors will later be classified by machine lear ning algorithms. The mathematical description of a parameter to classify is an e ssential part of the classification. If the mathematical description of a parameter does not correlate with the human lassification, the later numeric classification will not be satisfying. It is of advantage to generate a broad range of feature vectors out of which those with the highest descriptive power can be chosen afterwards. This can be done before the classification with a separate

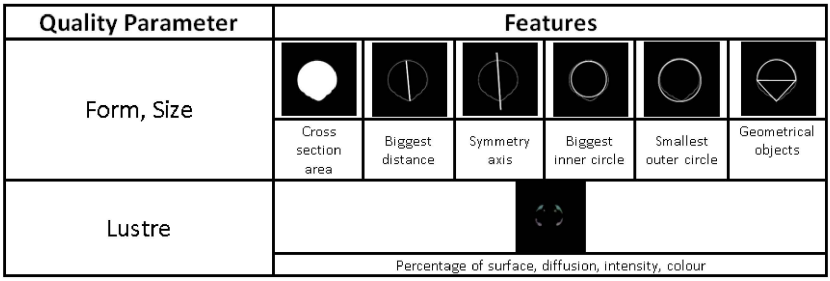


Fig. 6. Different possible features for the quality parameters form, size and lustre

correlation analysis (Principal component analysis, Self-Organizing Maps or linear discriminant analysis for example [9,10,11]) or with an iterative classification process, evaluating the classification results for different constellations of feature vectors. Several examples for features can be seen in Fig. 6.

4 Classification

To classify the quality of a pearl each quality parameter will be classified independently. This is due to the fact that the parameters are of different and independent character. For the classification of the size and the amount of defects no advanced machine learning algorithm is necessary. These two parameters are objectively measurable and can be directly graded in classes defined by pearl experts or governmental guidelines. For the parameters form, lustre and colour instead more effort has to be done, since classes defined by humans can only be described by subjective criteria. A key role plays the classification of the pearls colour. It is a crucial parameter for the pearls quality but colour classification is still an open problem in the domain of colour image processing. If an image is in RGB or related formats, colour object classification is a three dimensional problem. In our previous work [12] we reduced the classification problem to a two dimensional one. In the cited work we integrated the Hue-Saturation model in the RGB colour space. A separation of achromatic from chromatic values directly in the RGB space was done. The separation bases upon a projection of RGB vectors along a line through origin and the original vectors to a two dimensional plane with the achromatic line as normal vector (Fig. 7). The projection contains all chromatic information based on the definition of the HSV (Hue, Saturation and Value) space, while supressing achromatic information.

The projection furthermore corresponds to the linear character of the RGB space. It means a line separating the projection in two parts d escribes a certain proportional relation of RGB values to each other and therefore corresponds to the additive character of the RGB space. For an auto matic classification of colour based on this theory, an approach with linear SVM (Support Vector Machine) seems promising [13,14,15]. The SVM can be used to linearly separate between two pearls of different colour, by dividing the projection in two parts.

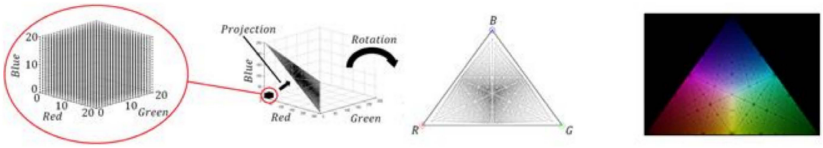


Fig. 7. A projection of RGB vectors to a plane with the achromatic line as normal to separate achromatic from chromatic information in the RGB space (left two images). The projection is afterwards turned into screen (third image). On the right side a coloured representation of the projection containing additional projection points can be seen

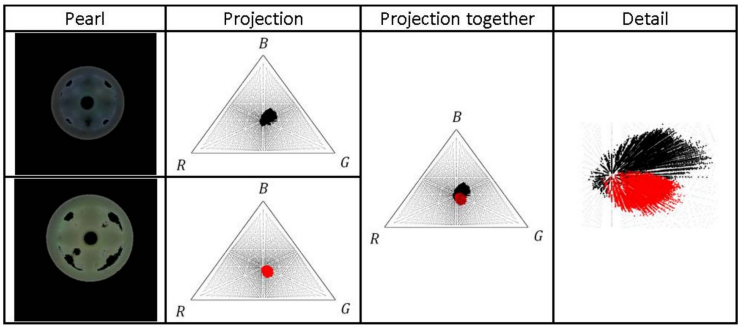


Fig. 8. On the left side two different segmented pearls are shown. In the second row the RGB vectors of each image are projected to suppress achromatic information. The two images on the right show the projection points of both images together

This linear separation correlates with the human biological processing of colour, as the projection is the barycentric representation of the RGB space. It corresponds as well to the human perception of colour, as the projection integrates the perceptual HSV model in the RGB space.

In Fig. 8 two different segmented pearls are shown. On the right side of each pearl the projection of the RGB vectors of the pearls describing the chromatic i nformation in the linear RGB geometry can be seen. Due to the fact that several projection points of the two pearls overlap, a soft-margin SVM seems an adequate choice. Our current analysis deals with the implementation of soft-margin linear SVM to identify colour vectors that uniquely describe a certain colour of a pearl. According to these results the goal is to implement an artificial reproduction of the human classification of the pearls colour.

5 Conclusion

In the present paper we introduced our ongoing work to implement an automatic classification of Tahitian pearls. A general methodology to classify pearls out of X-Ray and colour images was given. In the second section a new idea for greylevel edge detection was sketched and first promising results were shown. In the third

part possibilities to describe quality parameters gained out of colour images were discussed. In the fourth section an idea of colour classification based on our previous work on colour space theory was explained. Our current work deals with further developing the greylevel edge detection algorithm, to avoid global thresholding. An analysis of different machine learning algorithms to classify the calculated feature vectors to obtain a classification of the pearls quality is part of our current work as well.

References

1. http://www.peche.pf
2. http://wwz.ifremer.fr/cop
3. https://www.youtube.com/watch?v=JvKBSK6jbgw
4. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence (6), 679–698 (1986)
5. Prewitt, J.M.: Object enhancement and extraction. Picture Processing and Psychopictorics 10(1), 15–19 (1970)
6. Sobel, I., Feldman, G.: A 3×3 isotropic gradient operator for image processing. A talk at the Stanford Artificial Project in, pp. 271–272 (1968)
7. Senthilkumaran, N., Rajesh, R.: Edge detection techniques for image segmentation-a survey of soft computing approaches. International Journal of Recent Trends in Engineering 1(2), 250–254 (2009)
8. Jahne, B.: Digital Image Processing, vol. 5. Springer, Berlin (2002)
9. Nashat, S., Abdullah, M.Z.: Multi-class colour inspection of baked foods featuring support vector machine and Wilk's lambda analysis. Journal of Food Engineering 101(4), 370–380 (2010)
10. Laaksonen, J.T., Markus Koskela, J., Oja, E.: Class distributions on SOM surfaces for feature extraction and object retrieval. Neural Networks 17(8), 1121–1133 (2004)
11. Tobin, K.W., Bhaduri, B.L., Bright, E.A., Cheriyadat, A., Karnowski, T.P., Palathingal, P.J., Price, J.R.: Automated feature generation in large-scale geospatial libraries for content-based indexing. Photogrammetric Engineering and Remote Sensing 72(5), 531 (2006)
12. Loesdau, M., Chabrier, S., Gabillon, A.: Hue and Saturation in the RGB Color Space. In: Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D. (eds.) ICISP 2014. LNCS, vol. 8509, pp. 203–212. Springer, Heidelberg (2014)
13. Vapnik, V.N., Chervonenkis, A.J.: Theory of pattern recognition (1974)
14. Clarke, B.S., Fokoué, E., Zhang, H.H.: Principles and theory for data mining and machine learning, pp. 312–316. Springer, New York (2009)
15. Fang, Y., Pan, C., Liu, L., Fang, L.: Fast training of SVM via morphological clustering for color image segmentation. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005, Part I. LNCS, vol. 3644, pp. 263–271. Springer, Heidelberg (2005)

Noise Reduction in Ultrasound Images Based on the Concept of Local Neighborhood Exploration

Krystyna Malik, Bernadetta Machala, and Bogdan Smołka

Silesian University of Technology, Gliwice, Poland
{krystyna.malik,bernadetta.machala,bogdan.smolka}@polsl.pl

Abstract. In this paper a novel approach to the problem of speckle noise suppression in ultrasound images is presented. The described method is a modification of the bilateral denoising scheme and is based on the concept of local neighborhood exploration. The proposed filtering design, like the bilateral filter, takes into account the similarity of pixels intensities together with their spatial distance, and the filter output is calculated as a weighted average of the pixels belonging to the filtering window. The weights assigned to the pixels are determined by minimum connection costs of digital paths joining the central pixel of the filtering window and its neighbors. The comparison with existing denoising schemes shows that the new technique yields significantly better results in case of ultrasound images contaminated by multiplicative noise.

1 Introduction

Ultrasound imaging is one of the most widely used method in medical diagnostics, as it is non-invasive, has no known long-term side effects, is less expensive and simpler to use than other medical imaging techniques. However, the quality of a medical ultrasonic image is often degraded by the speckle noise, which is a random granular pattern produced mainly by multiplicative disturbances [1,2,3]. The speckle noise strongly impedes the visual evaluation of ultrasound images, affects edges and fine details and decreases the diagnostic value of ultrasound images. Speckle noise can also mask small, but diagnostically significant image features, and it reduces the efficiency of detection and recognition of the anatomical structures in medical images.

Therefore, in medical ultrasound image processing, the suppression of speckle noise, while preserving edges and image details, plays a crucial role for the diagnosis. Through the years, many techniques have been proposed for effective reduction of speckles in ultrasound images. The current approaches fall into three categories, including adaptive local filters, anisotropic diffusion based techniques and wavelet filters [2,4,3,5].

In this paper a new approach to multiplicative noise reduction in ultrasound images is presented. The new method is a based on the concept of the standard bilateral filter [6], whose modification was proposed in [7]. In that work, the described algorithm was applied to noise removal in color images contaminated by mixed Gaussian and impulsive noise. In the current manuscript, we propose

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_13

to use the elaborated framework to improve the quality of ultrasound images. The performed research showed that this method can be used to suppress the multiplicative noise in gray scale images and can be successfully employed for the enhancement of ultrasound images.

2 Proposed Noise Reduction Method

The proposed Geodesic Bilateral Filter (GBF), similarly to the bilateral filter [6], changes the value of each pixel by a weighted average of the intensity I of the samples in the filtering window W. In both cases, the new value of the pixel is calculated as:

$$\widetilde{J}(\boldsymbol{x}) = \frac{1}{Z} \sum\nolimits_{\boldsymbol{y} \in \mathcal{N}_{\boldsymbol{x}}} w(\boldsymbol{x}, \boldsymbol{y}) I(\boldsymbol{y}), \quad Z = \sum\nolimits_{\boldsymbol{y} \in \mathcal{N}_{\boldsymbol{x}}} w(\boldsymbol{x}, \boldsymbol{y}), \tag{1}$$

where $\mathcal{N}_{\boldsymbol{x}}$ is the local neighborhood of $\boldsymbol{x}$ and $w(\boldsymbol{x}, \boldsymbol{y})$ is the weight assigned to pixel at location $\boldsymbol{y}$, which belongs to $\mathcal{N}_{\boldsymbol{x}}$. In the bilateral filter, the weight is a result of multiplication of two components $w_S(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\|\boldsymbol{x} - \boldsymbol{y}\|^2 / 2\sigma_S^2\right)$ and $w_I(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\|I(\boldsymbol{x}) - I(\boldsymbol{y})\|^2 / 2\sigma_I^2\right)$, where σ_S and σ_I are filter parameters.

In case of the proposed Geodesic Bilateral Filter, for the calculation of the weights assigned to each pixel from the filtering window W, the image is treated as a graph and we utilize the Dijkstra's algorithm for finding the optimal connections between the pixels.

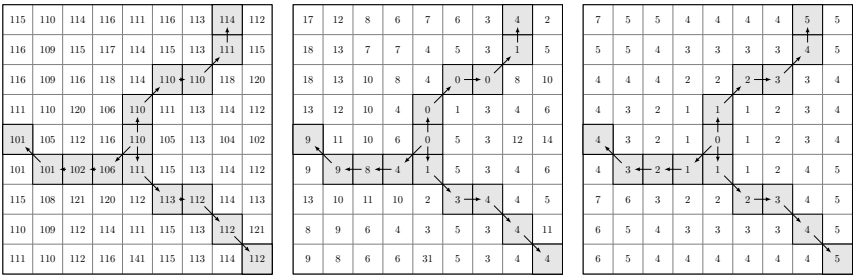
The connection cost of a pixel at position $\boldsymbol{y}$ is defined as a minimum sum of absolute differences between the pixels constituting a digital path connecting this pixel with the central pixel at $\boldsymbol{x}$. The connection cost is used to calculate a weight which is assigned to each pixel from the filtering window. In our approach, the weight function is calculated using only those pixels for which the length of the minimum crossing path is at least n, which can be treated as a filter parameter.

For the calculation of the connection costs in the analyzed neighborhood an array C, which includes the crossing cost between the central pixel and its neighbors, is created. In our method, it is necessary to introduce an additional array S, which includes the number of steps needed to join the central pixel with each sample belonging to W. Initially $\boldsymbol{S}(y) = 0$ for all pixels and this value is being changed during the creation of the path of the lowest cost.

Afterwards, the algorithm assigns to each pixel in the window the lowest connection cost with the central pixel. Finally, the array C includes the shortest distances for all the pixels in W and the array S includes the number of steps needed to create a digital path connecting the pixels in the local neighborhood as depicted in Fig. 1.

The weights are assigned only to those pixels from W, whose minimum digital path is of length not smaller than n. Thus, the weights of the pixels taken into the averaging is given by:

$$w(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-C(\boldsymbol{x}, \boldsymbol{y})^2 / h^2\right), \text{ for } (\boldsymbol{x}, \boldsymbol{y}) : S(\boldsymbol{x}, \boldsymbol{y}) \geq n, \tag{2}$$



(a) Filtering widow (b) Cost array C (c) Step array S

Fig. 1. Connection costs with some exemplary optimal paths

where n is a number of steps and $S(\boldsymbol{x}, \boldsymbol{y})$ is step function, $C(\boldsymbol{x}, \boldsymbol{y})$ is the cost function of the optimal path connecting $\boldsymbol{x}$ and $\boldsymbol{y}$ and is given by $C(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{m} |I(\boldsymbol{x}_i) - I(\boldsymbol{x}_{i-1})|$, where $\boldsymbol{x_0} = \boldsymbol{x}$ is the starting point of a path, $\boldsymbol{x_m} = \boldsymbol{y}$ and m is the number of path segments. The weight of pixels, for which the length of the minimum path is smaller than n are set to 0, which means that they are omitted.

3 Experiments

In this Section the Geodesic Bilateral Filter (GBF) is compared with the standard bilateral filter and other speckle noise reduction techniques in terms of objective quality measures. In the experiments we applied the restriction on the length of the digital path so that it is at least of length $n = 1$ (all pixels are considered) or $n = 2$, which is denoted as GBF and GBF^2 respectively. The effectiveness of the new filter was tested on the gray test images depicted in Fig. 2, whose intensities are in the range $[0, 1]$, contaminated by two types of multiplicative noise with standard deviations ranging from 0.1 to 0.3. In the first model, denoted as MG, the Gaussian distribution is used. The noisy pixels $I(\boldsymbol{x})$ are defined as $I(\boldsymbol{x}) = O(\boldsymbol{x}) + O(\boldsymbol{x}) \cdot \delta = O(\boldsymbol{x})(1 + \delta)$, where $O(\boldsymbol{x})$ is the undisturbed pixel intensity, δ is the Gaussian distributed random variable. The structure of the second model, denoted as MU, differs only in the definition of the variable δ, which is uniformly distributed.

The effectiveness of the new filtering design was compared with 13 different speckle reduction algorithm. We used following filters: Bilateral filter (BF), Median filter (MF), Lee statistic filter (LSF) [8], Kuan filter (KF) [8], Frost filter (FF) [9], Wiener filter (WF) [10], Butterworth low pass filter (BLPF), Anisotropic Diffusion (AD) [11], Speckle Reduction Anisotropic Diffusion (SRAD) [12], (VS) [13], SureShrink (SS) [14], BayesShrink (BS) [15]. To quantitatively evaluate the denoising methods we used the Peak Signal to Noise Ratio (PSNR) [16] and Mean Structural Similarity Quality Index (MSSIM) [17].

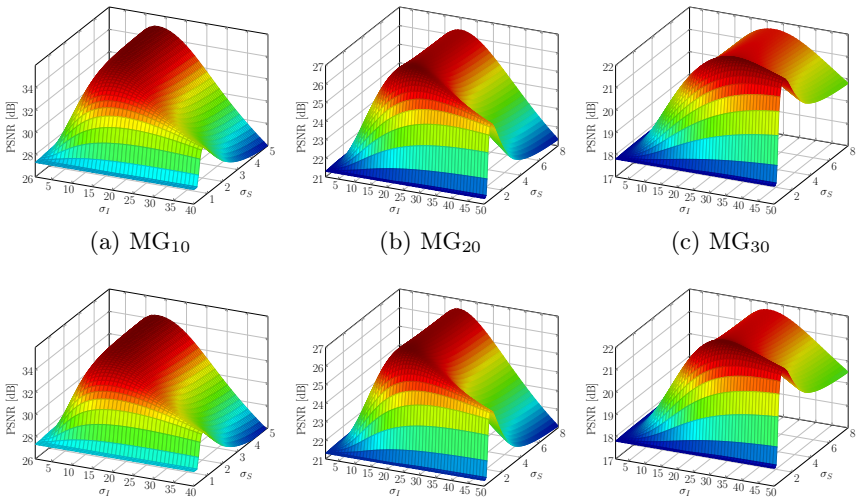
First, the relationship between the control parameters of the filters and the noise level was analyzed. As can be derived from (1) the properties of the BF

(a) TEST 1 (b) TEST 2 (c) TEST 3

Fig. 2. Gray test images used in the experiments

are controlled by the parameters σ_S and σ_I. Figure 3 shows the dependence of the PSNR on the σ_S and σ_I values for the noisy images restored by the bilateral filter.

The values of PSNR depend significantly on the σ_S and σ_I parameters. As can be observed the optimal values of σ_S and σ_I are noise level dependent for both boise models.



(a) MG_{10} (b) MG_{20} (c) MG_{30}

(d) MU_{10} (e) MU_{20} (f) MU_{30}

Fig. 3. Dependence of the PSNR on σ_S and σ_I for BF. Image TEST 2 was corrupted by MG (a, b, c) and MU noise (d, e, f).

Noisy test images were also restored by GBF and GBF^2. The filters were applied for different values of the parameter h in (2), and the dependence of PSNR on this parameter is depicted in Figs. 4 and 5. Examining the plots, it can be observed that optimal results depend on the tuning parameter h, which increases with the noise level. The obtained results show that the optimal h

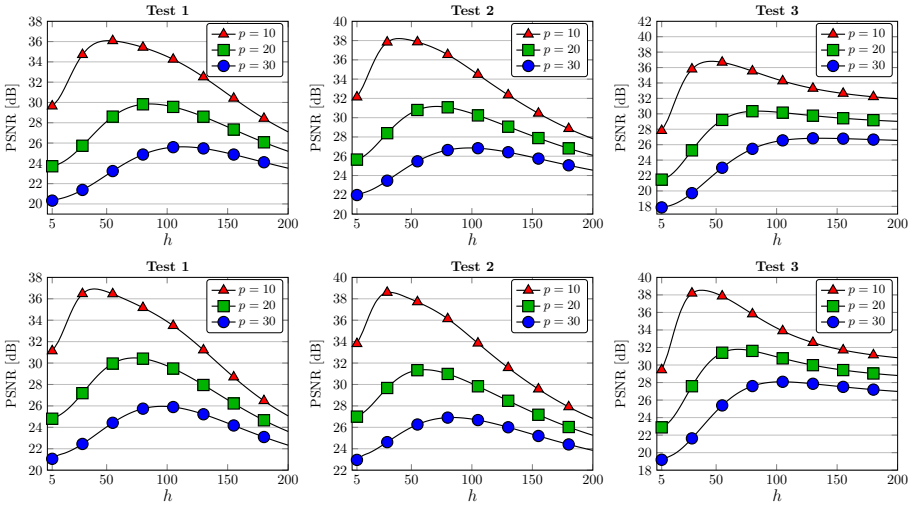


Fig. 4. Dependence of PSNR when applying GBF and GBF^2 on the h parameter for images corrupted with MG noise

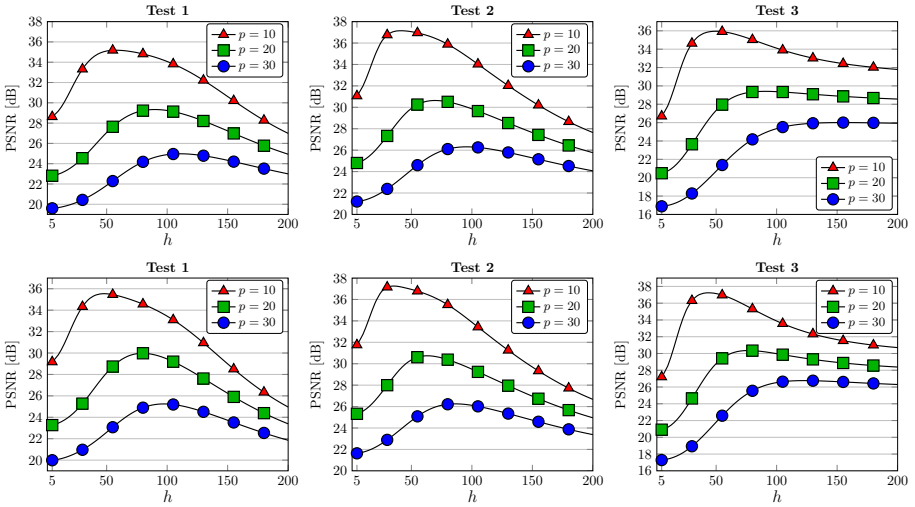


Fig. 5. Dependence of PSNR when applying GBF and GBF^2 on the h parameter for images corrupted with MU noise

parameter does not depend significantly on the image structure and the noise model.

The comparison of the efficiency of the proposed method with the other speckle reduction techniques are summarized in Tab. 1 and 2. The control parameters were selected experimentally to obtain optimal results in terms of the PSNR quality coefficient.

Table 1. Comparison of PSNR values of images contaminated by multiplicative noise (MG, MU) restored by the GBF and competing techniques

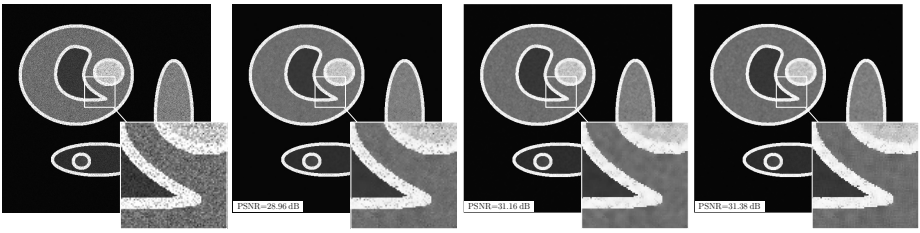
Images	Noise	Method													
		GBF	GBF^2	BF	MF	LSF	KF	FF	WF	BLPF	AD	SRAD	VS	SS	BS
Test1	MG_{10}	36.09	**36.92**	35.41	30.64	23.22	23.11	23.21	33.11	28.02	28.41	24.00	29.48	30.18	28.03
	MG_{20}	29.85	**30.47**	26.77	26.67	22.41	22.31	22.08	26.31	24.61	24.41	23.43	24.23	25.29	23.55
	MG_{30}	25.63	**25.96**	21.66	23.78	21.38	21.33	20.76	22.22	22.45	22.37	22.75	21.22	22.57	20.68
	MU_{10}	35.18	**35.53**	34.88	29.89	23.20	23.09	23.20	33.08	27.94	28.33	23.81	29.42	30.13	27.99
	MU_{20}	29.34	**29.98**	26.38	25.55	22.34	22.26	22.02	26.37	24.52	24.29	23.18	23.97	25.18	23.41
	MU_{30}	24.98	**25.24**	21.35	22.48	21.22	21.17	20.61	22.21	22.15	22.16	22.50	21.07	22.41	20.58
Test2	MG_{10}	38.21	**38.58**	37.73	34.78	24.94	24.81	23.77	34.58	29.77	30.19	26.28	31.56	23.49	30.02
	MG_{20}	31.16	**31.38**	28.96	28.96	24.01	23.88	23.58	27.74	26.14	25.99	25.71	25.91	26.90	25.31
	MG_{30}	26.86	**26.91**	23.61	25.68	22.87	22.77	22.17	23.45	23.83	23.72	24.69	22.63	23.98	22.26
	MU_{10}	37.15	**37.25**	37.14	33.62	24.90	24.73	24.74	34.44	29.67	30.09	25.88	31.49	23.45	29.94
	MU_{20}	30.64	**30.73**	29.01	27.72	23.93	23.82	23.52	27.90	26.04	25.90	25.19	25.88	26.84	25.42
	MU_{30}	26.31	**26.25**	23.45	24.40	22.73	22.63	22.07	23.54	23.65	23.54	24.41	22.58	23.91	22.24
Test3	MG_{10}	36.81	**38.52**	36.80	33.33	31.54	30.81	29.73	33.04	31.71	31.81	33.86	32.08	31.83	31.75
	MG_{20}	30.35	**31.78**	27.04	27.39	27.32	27.08	24.94	26.34	27.10	27.10	29.31	28.18	28.07	28.12
	MG_{30}	26.83	**28.08**	20.69	23.87	24.17	24.02	21.67	22.63	23.83	23.81	27.21	25.32	25.28	25.32
	MU_{10}	35.95	**37.23**	36.37	31.36	31.44	30.50	29.67	33.43	31.61	31.70	32.93	31.93	31.61	31.93
	MU_{20}	29.41	**30.35**	26.65	25.39	27.15	26.76	24.86	26.95	26.95	26.94	27.18	27.94	27.83	27.94
	MU_{30}	26.00	**26.76**	20.09	21.90	24.01	23.88	21.58	23.24	23.69	23.67	23.21	25.14	25.11	25.13

Table 2. Comparison of MSSIM values of images contaminated by multiplicative noise (MG, MU) restored by the GBF and competing techniques

Images	Noise	Method													
		GBF	GBF^2	BF	MF	LSF	KF	FF	WF	BLPF	AD	SRAD	VS	SS	BS
Test1	MG_{10}	0.965	**0.980**	0.970	0.936	0.936	0.935	0.932	0.946	0.913	0.911	0.867	0.916	0.935	0.898
	MG_{20}	0.914	**0.946**	0.890	0.863	0.873	0.872	0.865	0.866	0.836	0.835	0.941	0.845	0.878	0.825
	MG_{30}	0.873	**0.911**	0.830	0.813	0.824	0.824	0.815	0.812	0.798	0.802	0.925	0.796	0.834	0.776
	MU_{10}	0.960	**0.976**	0.968	0.913	0.935	0.934	0.930	0.946	0.911	0.908	0.708	0.915	0.934	0.897
	MU_{20}	0.902	**0.936**	0.880	0.834	0.869	0.868	0.861	0.864	0.833	0.832	0.774	0.840	0.874	0.821
	MU_{30}	0.854	**0.889**	0.820	0.781	0.821	0.820	0.812	0.810	0.789	0.799	0.869	0.794	0.834	0.774
Test2	MG_{10}	0.985	**0.986**	0.991	0.969	0.967	0.966	0.964	0.975	0.952	0.950	0.849	0.956	0.969	0.946
	MG_{20}	0.954	**0.959**	0.950	0.914	0.919	0.918	0.911	0.914	0.885	0.885	0.966	0.893	0.924	0.878
	MG_{30}	0.919	**0.926**	0.882	0.870	0.876	0.875	0.866	0.859	0.850	0.854	0.958	0.846	0.885	0.834
	MU_{10}	0.982	0.984	**0.991**	0.953	0.966	0.966	0.963	0.974	0.951	0.950	0.641	0.956	0.969	0.945
	MU_{20}	0.945	**0.951**	0.944	0.888	0.918	0.918	0.911	0.917	0.885	0.884	0.748	0.893	0.924	0.881
	MU_{30}	0.904	**0.910**	0.872	0.840	0.873	0.873	0.864	0.857	0.847	0.851	0.892	0.844	0.883	0.832
Test3	MG_{10}	0.936	**0.971**	0.957	0.829	0.862	0.856	0.837	0.850	0.827	0.843	0.967	0.889	0.895	0.888
	MG_{20}	0.790	0.889	0.714	0.578	0.637	0.635	0.592	0.585	0.577	0.602	**0.977**	0.707	0.715	0.707
	MG_{30}	0.644	0.783	0.418	0.407	0.462	0.461	0.420	0.408	0.406	0.428	0.848	0.539	0.549	0.539
	MU_{10}	0.928	**0.965**	0.955	0.751	0.862	0.855	0.837	0.860	0.827	0.843	0.937	0.889	0.895	0.890
	MU_{20}	0.761	0.869	0.688	0.472	0.637	0.634	0.593	0.605	0.578	0.603	**0.952**	0.703	0.712	0.703
	MU_{30}	0.623	0.759	0.388	0.319	0.464	0.463	0.420	0.425	0.407	0.430	**0.908**	0.548	0.558	0.548

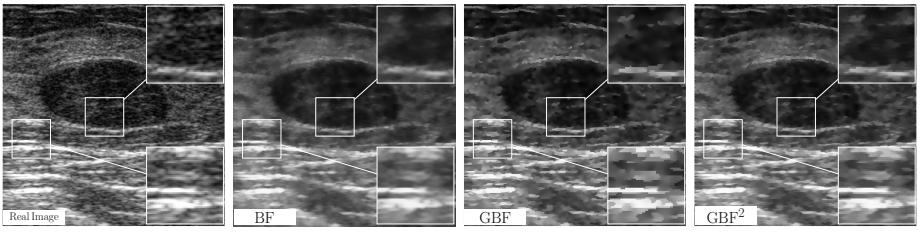
As can be observed, the results obtained for images restored using GBF and GBF^2 are significantly better in comparison with other denoising methods, and results obtained using GBF^2 are significantly better than when using the GBF.

Figures 6 and 7 summarize the restoration results of the new method and BF using synthetic and a real ultrasound image. As can be observed the filtering output is visually of better quality. The noise is suppressed, the edges are well preserved and important details are retained.



(a) Noisy image (b) BF (c) GBF (d) GBF^2

Fig. 6. Comparison of the efficiency of BF with the geodesic approaches



(a) Noisy image (b) BF (c) GBF (d) GBF^2

Fig. 7. Comparison of the efficiency of BF with the new filtering designs

4 Conclusions

The results of the performed experiments confirm the high efficiency of the proposed algorithm. The new filtering method yields significantly better results in comparison with other denoising techniques both in terms of objective quality measures and subjective fidelity. The introduction of minimal path length constraint improved significantly the effectiveness of the proposed filtering frmaework. The beneficial feature of the GBF and GBF^2 is the removal of noise with preservation of edges and image details. Additionally, the proposed method is easier to control as the BF properties are controlled by two parameters, while our filter needs only one tuning coefficient.

Acknowledgment. This work has been supported by grant BK-214/rau1/2013 of the Silesian University of Technology and was performed using the infrastructure supported by POIG.02.03.01-24-099/13 grant: GCONiI - Upper-Silesian Center for Scientific Computation.

References

1. Goodman, J.W.: Some fundamental properties of speckle. J. Opt. Soc. Am. 66(11), 1145–1150 (1976)
2. Michailovich, O., Tannenbaum, A.: Despeckling of medical ultrasound images. IEEE Trans. on Ultrasonic, Ferroelectic, Frequence Control 53, 64–78 (2006)
3. Sarode, M., Deshmukh, P.: Reduction of speckle noise and image enhancement of images using filtering technique. International Journal of Advancements in Technology 2(1), 30–38 (2011)
4. Loizou, C.P., Pattichis, C.S.: Despeckle Filtering Algorithms and Software for Ultrasound Imaging. Synthesis Lectures on Algorithms and Software in Engineering. Morgan and Claypool Publishers (2008)
5. Rosa, R., Monteiro, F.C.: Speckle ultrasound image filtering: Performance analysis and comparison. In: Tavares, J.M.R.S., Natal, J.R.M. (eds.) Computational Vision and Medical Image Processing IV, VIPIMAGE 2013, pp. 65–70 (2013)
6. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proceedings of the Sixth International Conference on Computer Vision, ICCV 1998, p. 839. IEEE Computer Society, Washington, DC (1998)
7. Malik, K., Smolka, B.: Improved bilateral filtering scheme for noise removal in color images. In: The International Conference on Informatics and Applications (ICIA 2012) (2012)
8. Kuan, D.T., Sawchuk, A., Strand, T.C., Chavel, P.: Adaptive restoration of images with speckle. IEEE Transactions on Speckle Ultrasound Image Acoustics, Speech and Signal Processing 35(3), 373–383 (1987)
9. Frost, V.S., Stiles, J.A., Shanmugan, K.S., Holtzman, J.: A model for radar images and its application to adaptive digital filtering of multiplicative noise. IEEE Trans. on PAMI 4(2), 157–166 (1982)
10. Jin, F., Fieguth, P., Winger, L., Jernigan, E.: Adaptive wiener filtering of noisy images and image sequences. In: International Conference on Image Processing, ICIP 2003, vol. 3, pp. 349–352 (2003)
11. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. on PAMI 12, 629–639 (1990)
12. Yu, Y., Acton, S.T.: Speckle reducing anisotropic diffusion. IEEE Trans. on Image Processing 11(11), 1260–1270 (2002)
13. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455 (1994)
14. Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association, 1200–1224 (1995)
15. Chang, S.G., Yu, B., Vetterli, M.: Adaptive wavelet thresholding for image denoising and compression. IEEE Trans. on Image Processing 9(9), 1532–1546 (2000)
16. Prakash, K.B., Babu, R.V., VenuGopal, B.: Image independent filter for removal of speckle noise. International J. of Computer Science Issues 8(5), 196–200 (2011)
17. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error measurement to structural similarity. IEEE Trans. on Image Processing 13, 600–612 (2004)

Viterbi Algorithm for Noise Line Following Robots

Przemysław Mazurek

Department of Signal Processing and Multimedia Engineering,
West–Pomeranian University of Technology, Szczecin,
26. Kwietnia 10 St., 71126 Szczecin, Poland
przemyslaw.mazurek@zut.edu.pl

Abstract. Image processing and tracking of noise line is considered in this paper. Such line could be obtained for selected application where the line is not direct, but obtained from the image content. The estimation of line allows control of line following robot. Local 2D filter based on standard deviation estimator is applied for the preprocessing of the image. The Viterbi algorithm is applied for the line tracking using assumed Markov model of line. Monte Carlo approach is used for the estimation of the tracking system performance.

1 Introduction

Tracking system are applied in numerous applications and most of them assume Detection and Tracking scheme [3]. The input image is thresholded using fixed or adaptive threshold level and the binary image is obtained. The tracking algorithm adds robustness to estimated trajectory (line) under noisy conditions. There are numerous of tracking algorithms and the Kalman filter is typically applied.

Low quality images are difficult to process due to high noise, because optimal threshold level cannot be determined. Alternative tracking scheme should be applied – TBD (Track–Before–Detect). There are a few groups of TBD algorithms [3,18] and the Viterbi algorithm is assumed [20] in this paper. The tracking is applied for the estimation of cumulative value for trajectory without prior thresholding. TBD algorithms allow raw signal processing, that gives superior performance in comparison with Detection and Tracking scheme based tracking systems.

Specific input image case is related to the images with noise only. It means that the background object trajectory is noise only. Additional preprocessing is necessary, because TBD algorithms assume positive signal of the object, and zero mean of the background noise.

1.1 Related Works

Line following robots are well known and applied in manufactures [5], especially. Typical line following robot application assumes high contrast between the line

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_14

and the background. Such assumption allows the application of simple optical systems and simple algorithms [7]. The line could be deteriorated and lighting conditions could be variable. Locally variable lighting conditions, and noises related to the deteriorated background and real line, as well as false lines, are the sources of low quality of acquired image. Positive signal of line is assumed and dark background in most cases, but there are applications in which the line is not observed directly. The line could be an edge between two areas or very low quality line related to the real environment. Such cases are typical for the applications [16] like harvesting, trash compacting, snow or sand plowing on airport runways and many more. Road lane estimation is the similar task for automatic control of vehicle or as Lane Departure Warning system. Image processing and pattern recognition techniques are necessary for such local navigation, because GPS systems have low precision for spatial location (about a few meters) for moving platforms

Hough Transform approach for highly deteriorated line is considered in [19] for the road lane estimation as well as in agricultural application [1]. The application of LIDAR for the acquisition of 3D structures for line estimation is considered in [21]. Monte Carlo approach for line estimation is proposed in [15].

The application of the preprocessing using local standard deviation is proposed in [10] and the application of maximal autocovariance is described in [14] for different kind of TBD algorithm – Spatio–Temporal TBD [9].

1.2 Content and Contribution of the Paper

In this paper the Viterbi algorithm is applied for the noise line tracking and the performance of such system is analyzed using Monte Carlo approach, that allows testing many uncorrelated scenarios. The Viterbi algorithm is introduced briefly in Section 2. Noise line preprocessing is considered in Section 3. Monte Carlo approach is applied and the results for different configurations and are shown in Section 4. Discussion related to the obtained results is provided in Section 5. The final conclusions are in Section 6.

2 The Viterbi Algorithm

The Viterbi algorithm is one of the dynamic programming algorithms [2]. It uses trellis with a set of nodes, and in this paper nodes are assigned to the pixels' grid. The computation of the trajectory (path) is based on the calculation of maximal value for paths allowed by Markov transition model [12]. Markov model is related to the trajectory shape, allowed by the trellis geometry and include probabilities of transition between nodes. In this paper it is assumed that the line is one direction starting from the bottom row of the image toward top row. Many applications requires estimation the most valuable bottom part of image – nearest to robot. Robots control is based on the this area, and top part of image could be significantly deteriorated due perspective camera view.

Image analysis starts from first row $n = 1$ and zero values V are assigned to the nodes:

$$V_{n=1},. = 0, \tag{1}$$

where V_n, x is the largest value assigned to the node n, x and the coordinates correspond to image coordinates. The assumed trellis is shown in Fig. 1.

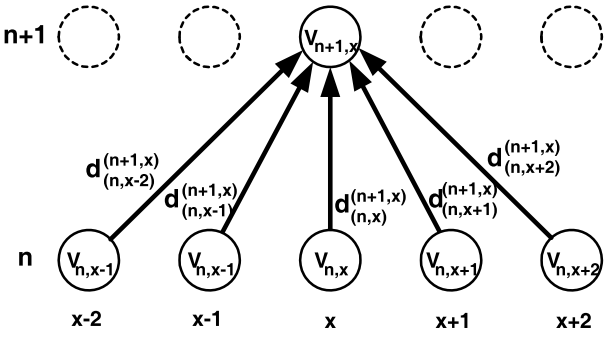


Fig. 1. Local paths in example trellis

In this paper local cost $d_{n,x+g}$ is related to the pixel value. The next row of values $n = 2$ is computed using the following formula, that uses local cost d:

$$V_{n+1,x} = \max\left(V_{n,x+g} + d_{n,x+g}^{n+1,x}\right). \tag{2}$$

for set of transitions:

$$g \in \{-2, -1, 0, +1, +2\} \tag{3}$$

The selection of the best local path to the particular node is preserved additionally:

$$L_n^{n+1,x} = \arg\max_g \left(V_{n,x+g} + d_{n,x+g}^{n+1,x}\right), \tag{4}$$

where L is local transition. The projection of values from first row is processed up to selected arbitrary n_{max} row. After reaching of this row the first phase (forward phase) of the Viterbi algorithm is achieved. The solution is the node with maximal value

$$P_{opt} = V_{n=n_{max},.}, \tag{5}$$

and the node number is:

$$x_{n=n+max} = \max_x \left(V_{n=n_{max},x}\right). \tag{6}$$

The second phase of the Viterbi algorithm (backward phase) is applied for the calculation of the best and first local transition. The obtained path of n_{max} length is rejected and only mentioned transition is valid.

The following recursive formula is applied for finding optimal solution:

$$x_{n-1} = x_n + L_{n-1}^{n,x} \tag{7}$$

for successively decremented row numbers:

$$n = n_{max}, \cdots, 2. \tag{8}$$

The transition between first and second row for x_1 point is the result of the Viterbi algorithm. The sliding window approach is applied in next step so overall process starts from next row with n_{max} depth analysis. The obtained set of points could be processed additionally using another algorithm for the trajectory filtering if it is desired.

3 Noise Line Preprocessing

The line is typically observed as a positive signal in the input image. Noise signal requires additional preprocessing and it could be obtained by the application of local filter that estimates noise parameter. As a local filter 2D local standard deviation $S(y, x)$ is applied:

$$S(y,x) = \sqrt{\frac{1}{(2x_N+1)(2y_N+1)} \sum_{\Delta y=-y_N}^{y_N} \sum_{\Delta x=-x_N}^{x_N} (\overline{X}_{y,x} - X_{y+\Delta y, x+\Delta x})^2}, \tag{9}$$

where X denote input image and $\overline{X}$ denote mean value for the corresponding area. The scaling coefficient $(2x_N+1)(2y_N+1)$ could be omitted for the efficient implementation. The window of the local estimation of standard deviation has $(2x_N+1) \times (2y_N+1)$ size. Overall system is shown in Fig. 2.

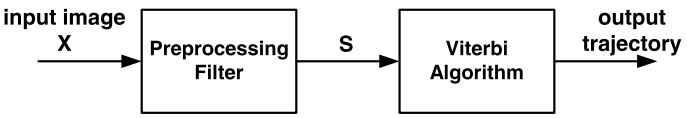


Fig. 2. Schematic of tracking system

4 Results

In Fig. 3 a single example tracking scenario is presented. The line is tracked from row 1 to 120, because the deep of analysis $n_{max} = 80$ is assumed. This line has standard deviation 1.0 and image is disturbed by a few lines with standard deviation selected by the random number generator from 0.2 to 1.0 range. All lines are disturbed by additive Gaussian noise and all lines are at the edge of visibility by human. The application of the filter that calculates local standard

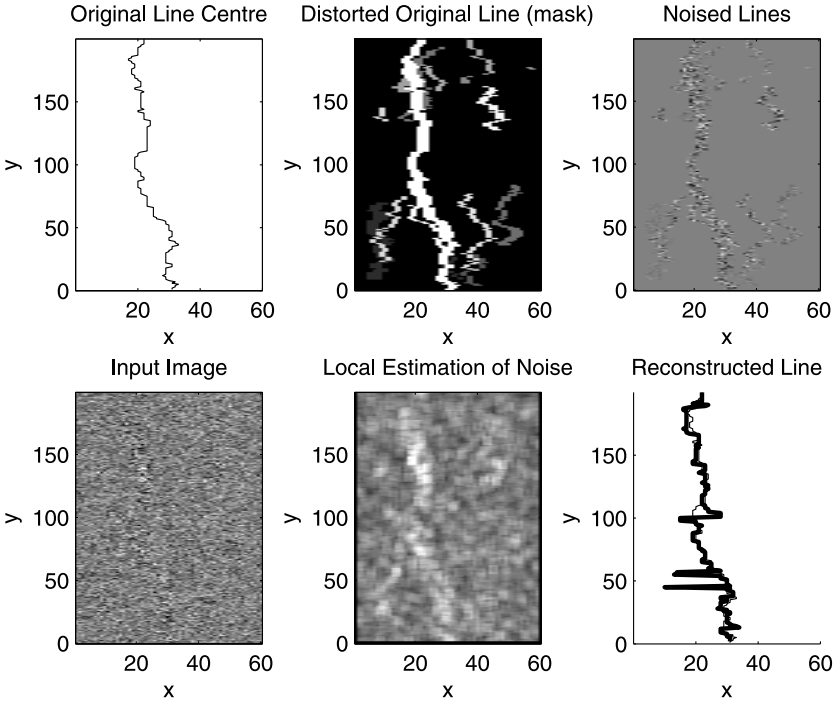


Fig. 3. Example tracking case

deviation allows the partial detection of the lines. The Viterbi algorithm gives the trajectory estimation that is comparable to the original.

The window size of the preprocessing filter is $H \times W = 5 \times 3$, where W and H are the width and height respectively.

Monte Carlo approach is applied for the analysis of the performance of proposed tracking system. This approach allows the testing of system before implementation for different scenarios.

In Fig. 4 the results for the different n_{max} values and for different noise values are shown.

5 Discussion

Four filters are compared (Fig. 4) using Monte Carlo approach and there are 300 cases in each configuration. Smooth line on figures means that test is sufficient. The window size of the filter influences the result. The filter with larger window height (11) gives better results. This is the result of the directional (vertical) type of line. The fitting of filter window to the direction of the line improves the detection. The width of filter window influences the results also. The width of line is variable, but it allocates a few pixel typically. Larger width of window reduces the tracking performance, because narrow windows are better fitted inside line.

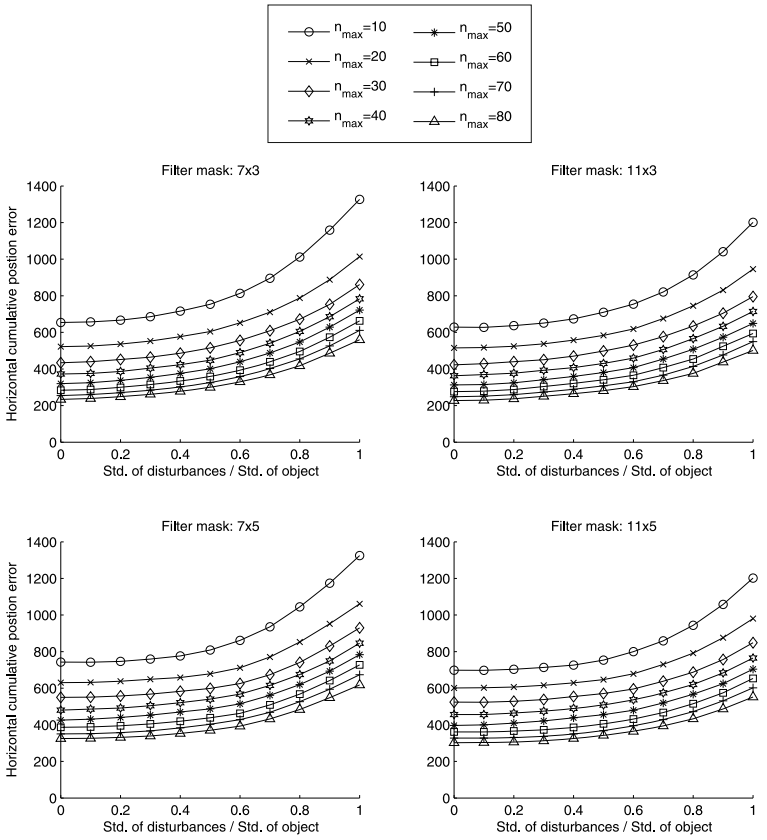


Fig. 4. Cumulative error for horizontal position for different filters and noise conditions

The influence of the deep of analysis is well visible (Fig. 4). Larger n_{max} values give better results, but the benefits of performance is reduced for $n_{max} > 30$. The selection of the proper n_{max} value depends on available computation power and the cost is linear.

6 Conclusions

The application of the Viterbi algorithm with preprocessing allows the noise line tracking. Assumed local filtration based on estimation of the standard deviation allows the conversion of image to the desired measurement space.

There are many TBD algorithms and similar techniques could be applied, e.g. Particle Filters algorithm [6].

Filtering may work also as TBD algorithm if is applied in the specific direction and it is a kind of the hierarchical preprocessing [11]. The application of additional processing between both algorithms is possible also for the emphasis of the lines, using techniques shown in [8,4] for example.

The variable line width could be processed by the application of the filter banks [13] for the selection of best filtering space, instead processing of image with fixed filtering mask.

The computation cost is low for both parts and image could be processed using modern processors or microcontroller. Large images could be processed using GPGPUs (General–Purpose Graphics Processor Units) in real–time [9,17]. There are the Viterbi algorithm accelerators that are available in specific DSPs (Digital Signal Processors). Parallel processing of both algorithms is possible also.

Acknowledgment. This work is supported by the UE EFRR ZPORR project Z/2.32/I/1.3.1/267 /05 "Szczecin University of Technology – Research and Education Center of Modern Multimedia Technologies" (Poland).

References

1. Astrand, B., Baerveldt, A.: A vision-based row-following system for agricultural field machinery. Mechatronics 15(2), 251–269 (2005)
2. Bertsekas, D.: Dynamic Programming and Optimal Control, vol. I. Athena Scientific (1995)
3. Blackman, S., Popoli, R.: Design and Analysis of Modern Tracking Systems. Artech House (1999)
4. Frejlichowski, D., Forczmański, P.: General shape analysis applied to stamps retrieval from scanned documents. In: Dicheva, D., Dochev, D. (eds.) AIMSA 2010. LNCS, vol. 6304, pp. 251–260. Springer, Heidelberg (2010)
5. Horan, B., Najdovski, Z., Black, T., Nahavandi, S., Crothers, P.: Oztug mobile robot for manufacturing transportation. In: IEEE International Conference on Systems, Man and Cybernetics (SMC 2011), pp. 3554–3560 (2011)
6. Huang, D., Xue, A., Guo, Y.: A particle filter track–before–detect algorithm for multi–radar system. Elektronika ir Elektrotechnika 19(5), 3–8 (2013)
7. Lech, P., Okarma, K.: Optimization of the fast image binarization method based on the monte carlo approach. Electronics and Electrical Engineering 20(4), 63–66 (2014)
8. Marchewka, A.: Crack detection on asphalt surface image using local minimum analysis. Advances in Intelligent and Soft Computing 84, 353–359 (2010)
9. Mazurek, P.: Optimization of bayesian track-before-detect algorithms for GPGPUs implementations. Electrical Review R 86(7), 187–189 (2010)
10. Mazurek, P.: Track-before-detect algorithm for noise objects. Measurement Automation and Monitoring 56(10), 1183–1185 (2010)
11. Mazurek, P.: Hierarchical track-before-detect algorithm for tracking of amplitude modulated signals. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 3. AISC, vol. 102, pp. 511–518. Springer, Heidelberg (2011)
12. Mazurek, P.: Code reordering using local random extraction and insertion (LREI) operator for GPGPU-based track-before-detect systems. Soft Computing 18(6), 1095–1106 (2013)
13. Mazurek, P.: Track–before–detect filter banks for noise object tracking. International Journal of Electronics and Telecommunications 59(4), 325–330 (2013)

14. Mazurek, P.: Preprocessing using maximal autocovariance for spatio-temporal track-before-detect algorithm. In: Choras, R.S. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 45–54. Springer, Heidelberg (2014)
15. Okarma, K., Lech, P.: A fast image analysis for the line tracking robots. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010, Part II. LNCS, vol. 6114, pp. 329–336. Springer, Heidelberg (2010)
16. Ollis, M.: Perception Algorithms for a Harvesting Robot. CMU-RI-TR-97-43, Carnegie Mellon University (1997)
17. Pietraszek, J., Szczotok, A., Kocyłowska, E.: Factorial approach to assessment of gpu computational efficiency in surogate models. Advanced Materials Research 874, 157–162 (2014)
18. Stone, L., Barlow, C., Corwin, T.: Bayesian Multiple Target Tracking. Artech House (1999)
19. Taubel, G., Yang, J.S.: A lane departure warning system based on the integration of the optical flow and Hough transform methods. In: 2013 10th IEEE International Conference on Control and Automation (ICCA), Hangzhou, China, June 12-14, pp. 1352–1357 (2013)
20. Viterbi, A.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory 13(2), 260–269 (1967)
21. Zhang, J., Chambers, A., Maeta, S., Bergerman, M., Singh, S.: 3d perception for accurate row following: Methodology and results. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, November 3-7, pp. 5306–5313 (2013)

Influence of Impulse Noise on ST–TBD and Viterbi Track–Before–Detect Algorithms

Przemysław Mazurek

West–Pomeranian University of Technology, Szczecin,
Department of Signal Processing and Multimedia Engineering,
26. Kwietnia 10 St., 71126 Szczecin, Poland
przemyslaw.mazurek@zut.edu.pl

Abstract. Two algorithms that are used for the Track–Before–Detect (TBD) applications are compared. The impulse noise gives different results for Spatio–Temporal TBD and the Viterbi TBD algorithms. The selection of algorithm should be based on expected impulse noise source. Two cases – impulse noise related to the maneuver and trajectory independent are compared in this paper. The noise suppression techniques could be applied for the reduction of introduced trajectory errors.

1 Introduction

Tracking systems are very important nowadays [2]. The majority of systems are applied in surveillance applications. Radar and infrared (vision–based) tracking systems are applied in the air and space surveillance. Many systems are used for military purposes, but civil systems are very important, also. There are a lot of algorithms proposed for such systems and two main approaches are applied. The first approach is based on the detection and tracking. The input signal is processed by the thresholding algorithm, with fixed or adaptive threshold, for the detection of the object position and removal of the background noise. Such case is rare, and in many applications there are much false detections, so another algorithm (tracking algorithm) is applied. Tracking algorithms use the motion model and the gate for the selection of observation area for further observations. The application of tracking algorithm greatly improves performance of the system and the Kalman, EKF, Bayes filters are applied typically [2,11]. The system for multi–object tracking is more sophisticated, because advanced track maintenance is necessary and appropriate assignment algorithm is desired.

The ratio of object's signal to the background noise (SNR) influences the tracking performance. The threshold level should be low for the detection of weak signals, but it increases greatly the number of false detections. Low SNR cases are very difficult to process using the detection and tracking approach. Such scenarios could be processed using TBD (Track–Before–Detect) approach only.

TBD systems uses the tracking of all possible trajectories. Signals are accumulated using trajectories so multidimensional filtering is obtained. Values for

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_15

trajectories are de–noised (calculating the mean) so the detection is not related to the noised signal samples, but trajectory values.

1.1 Related Works

There are a few classes of the TBD algorithms – ST–TBD algorithm and the Viterbi algorithm [1,4] that are considered in this paper. The problem of impulse noise for ST–TBD algorithm in considered in [7]. The solutions based on the saturation for ST–TBD are considered in [7].

1.2 Content and Contribution of the Paper

The problem of impulse noise for ST–TBD algorithm is briefly introduced in Section 2, based on the previous work [7]. In this section ST–TBD algorithm is also presented. The Viterbi algorithm is considered in Section 3 and the noise related behavior is analyzed in Section 4. Section 5 consists of discussion of four examples from Section 4 that illustrate different behaviors of impulse noise depending on the algorithm.

2 Spatio–Temporal TBD Algorithm

ST–TBD algorithm is a kind of the multidimensional recursive filter [6]. The following pseudocode shows processing details:

Start
// Initial:

$$P(k=0,s)=0 \tag{1}$$

For $k \geq 1$ and $s \in S$
// Motion Update:

$$P^{-}(k,s)=\int_{S} q_k(s|s_{k-1})P(k-1,s_{k-1})ds_{k-1} \tag{2}$$

// Information Update:

$$P(k,s)=\alpha P^{-}(k,s)+(1-\alpha)X(k) \tag{3}$$

EndFor
End
where:

S	– state space: e.g. 2D position and motion vectors,
s	– state (spatial and velocity components),
k	– time moment,
α	– smoothing coefficient $\alpha \in (0,1)$,
$X(k)$	– measurements,
$P(k,s)$	– estimated value of objects,
$P^{-}(k,s)$	– predicted value of objects,
$q_k(s\|s_{k-1})$	– state transitions (Markov matrix).

Input measurements are processed by the Information Update formula. This formula is the kind of the exponential filter and input measurements are mixed with the predictions. The predictions are calculated by the Motion Update formula using Markov matrix that describes motion model. There are two processes in ST–TBD algorithm. The Information Update formula is responsible for sharpening process and the Motion Update formula is responsible for blurring process. The blurring process is related to the dispersion of values by the predictor due to Markov matrix. The balance between blurring and sharpening is defined by the smoothing coefficient mainly.

The state–space of ST–TBD could be directly related to measurement space (input image). Two dimensions related to the input image are extended due to motion model. The third dimension could be a set of motion vectors. The output of the algorithm is the P or P^{-}.

The impulse noise could be analyzed using 1D input signal, instead of 2D, sufficiently. Simple Markov transition could be modeled by the following set of equations:

$$x(k+1, n-1) = a \cdot x(k, n), \tag{4}$$

$$x(k+1, n) = a \cdot x(k, n), \tag{5}$$

$$x(k+1, n+1) = a \cdot x(k, n), \tag{6}$$

where k is the time moment, n if the position, a is weight (transition probability) that fulfill the following requirement:

$$a + a + a = 1. \tag{7}$$

Vector related to response after K time could be calculated [3] using the following formula:

$$x_K = x_0 Q^K \tag{8}$$

where starting vector x_0 is defined as:

$$x_0 = [\cdots \; 0 \; 0 \; c \; 0 \; 0 \cdots], \tag{9}$$

and c is amplitude of the impulse and Q is Markov matrix. Markov matrix is sparse [7,8].

$$Q = \begin{bmatrix} \cdots & & & & & & & \\ \cdot\cdot & a & & & & & & \\ \cdot & a & a & & & & & \\ & a & a & a & & & & \\ & & a & a & a & & & \\ & & & a & a & a & & \\ & & & & a & a & \cdot & \\ & & & & & a & \cdot\cdot & \\ & & & & & & \cdots & \end{bmatrix} \tag{10}$$

The response of ST–TBD filter is, after K–steps could be calculated using the following formula:

$$P_K = (1-\alpha)\alpha^{(K-1)} x_0 Q^K. \tag{11}$$

This formula uses α coefficient that is part of the Information Update formula.

3 The Viterbi Algorithm

The Viterbi algorithm could be applied in TBD applications, also. This algorithm uses two processing phases. The simplest case of the Viterbi algorithm uses direct pixel–to–node assignment. The trajectory estimation is based on the analysis of the transitions between nodes for assumed trellis. Example part of trellis for 1D case is shown in Fig. 1. The transition cost could is fixed (a–value).

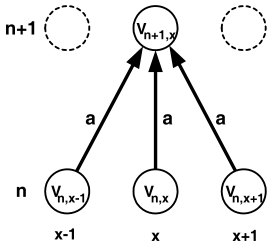


Fig. 1. Local paths in example trellis

The Viterbi algorithm estimates the transition between first and second row of nodes using multiple rows and in 1D case rows are 1D signals variable in time, where time corresponds to the node row number. There are n_{max} observations necessary for the computation of the state transition between first and second time moment (row). The Viterbi algorithm does not use obtained transition for $1, \cdots, n_{max}$ set during calculation of transition for next $2, \cdots, (n_{max}+1)$ set. Computation starts from the assignment of zero values to the total cost V for nodes of the first row ($n = 1$):

$$V_{n=1,.} = 0 \tag{12}$$

The local cost d is related to the pixel value X multiplied by weight a (that could be omitted, because is fixed for all local paths) and the next row is:

$$V_{n+1,x} = \max\left(V_{n,x+g} + X_{n,x+g}\right) \tag{13}$$

The transition is preserved in L using the following formula:

$$L_n^{n+1,x} = \arg\max_g \left(V_{n,x+g} + X_{n,x+g}\right). \tag{14}$$

The first phase of the Viterbi algorithm is computed up to n_{max}. The solution for the first phase (path) is the node with maximal value:

$$P_{opt} = \max\left(V_{n=n_{max},.}\right), \tag{15}$$

that is related to the single column:

$$x_{n=n_{max}} = \arg\max_x \left(V_{n=n_{max},x}\right). \tag{16}$$

The selection of best–valued node allows the starting the second phase, that is applied for the selection of most probable path over trellis using the following recursive formula:

$$x_{n-1} = x_m + L_{n-1}^{n,x}, \tag{17}$$

for successively decremented row numbers:

$$n = n_{max}, \cdots, 2. \tag{18}$$

The solution is the first transition and the first node row. This process is repeated using for another set of observations independently on previous results (e.g. $n = 2, \cdots, n_{max} + 1$). Such assumption has some disadvantages, from computational point–of–view and tracking continuity, but is important in the case of noised measurements.

4 Examples of Influence by Impulse Noise

Impulse noise is specific in tracking systems. Such noise could be observed in sensor due to properties of sensor or as physical phenomenon that influences the sensor. Vision systems are sensitive on high energy particles, that create false observations and such cases are typical in astronomy. The impulse that is not related to any existing trajectory could be detected and removed from measurement using simple threshold techniques. The most important problem is the impulse noise related to the true trajectory, especially for the low–SNR objects. An example of impulse noise is the signal peak from plane during the maneuver or from thumbing satellite. Such impulses could be used for the tracking improvement, but the 'SNR Paradox' [7] shows the real problem. The peak values incorporated to the TBD tracking preserves state of the object before maneuver. The signal after impulse is hidden by this peak so switching between trajectories is difficult.

In Fig. 2 the tracking example for ST–TBD algorithm is shown. The impulse is located at coordinates: $x = 30, y = 55$. The peak value is very high – 1000 times

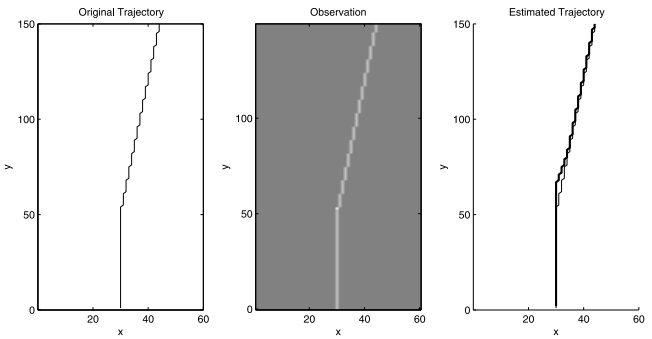


Fig. 2. Tracking example for ST–TBD algorithm – impulse noise related to trajectory

higher then signal level. No background noise is added for the simplification of the results. In Fig. 3 the example where the impulse is not related to the trajectory ($x = 10, y = 55$) for smoothing coefficient $\alpha = 0.98$ is shown. Both examples use the maximal value of the prediction as output.

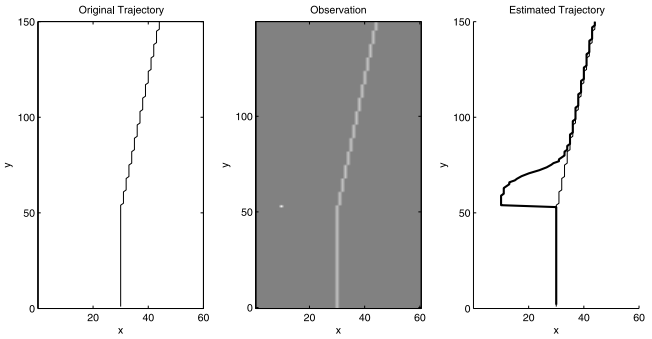


Fig. 3. Tracking example for ST–TBD algorithm – impulse noise not related to trajectory

The Viterbi algorithm for impulse related to the trajectory is shown example for in Fig. 4. The deep of analysis is $n_{max} = 25$. In Fig. 5 is shown example for the impulse located outside of trajectory.

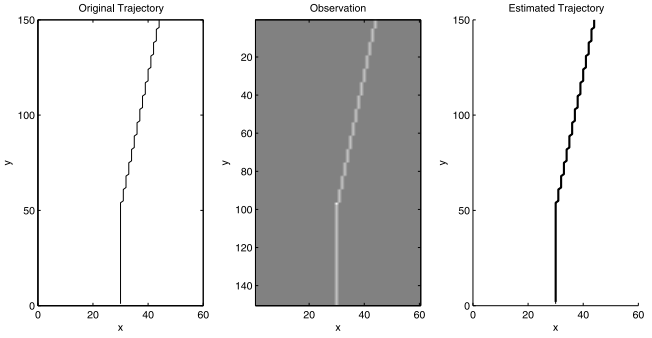


Fig. 4. Tracking example for the Viterbi algorithm – impulse noise related to trajectory

5 Discussion

The selection of proper TBD algorithm in the case of impulse noise depends on the situation. The ST–TBD algorithm preserves previous state if the impulse is related to the trajectory (Fig. 2) that is the effect of the impulse noise (11). The Viterbi algorithm has better properties because the considered impulse does not introduce tracking errors (Fig. 4). Impulse noise not related to the trajectory

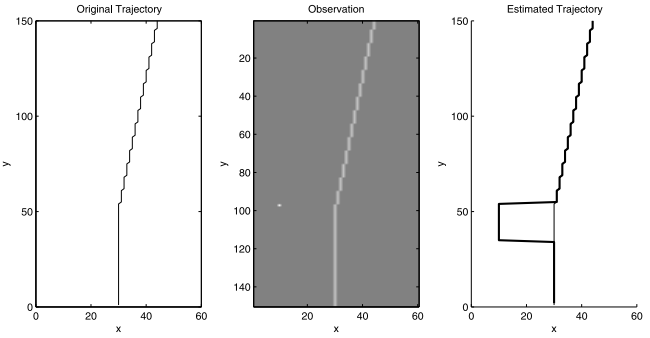


Fig. 5. Tracking example for the Viterbi algorithm – impulse noise not related to trajectory

is significant problem for the tracking and the position error is huge. ST–TBD algorithm gives the error that is decreased in time (y–axis). The Viterbi algorithm results depend on the implementation and in this example the position for $y = 35, \cdots, 55$ is parallel. The error is related to the time moments before peak occurs. The output of ST–TBD algorithm has introduced error after the peak.

6 Conclusions

Impulse noise is very important for TBD algorithms, because may introduce different kinds of errors. Such noise may improve or reduce tracking quality depending on algorithm and type of impulse (related or not to the object). Four illustrative examples are provided in this paper.

Additional trajectory filtering [5] may improve results independently on noise suppression in the spatial domain. There are applications where the noise is observed only, so object's signal is the noise pattern [9,10]. Such case needs preprocessing of measurements independently on impulse noise suppression.

Acknowledgment. This work is supported by the UE EFRR ZPORR project Z/2.32/I/1.3.1/267 /05 "Szczecin University of Technology – Research and Education Center of Modern Multimedia Technologies" (Poland).

References

1. Bertsekas, D.: Dynamic Programming and Optimal Control, vol. I. Athena Scientific (1995)
2. Blackman, S., Popoli, R.: Design and Analysis of Modern Tracking Systems. Artech House (1999)
3. Grinstead, C., Snell, J.: Introduction to Probability. AMS (1997)
4. Haykin, S., Moher, M.: Communication Systems. John Wiley & Sons (2009)

5. Marchewka, A.: Crack detection on asphalt surface image using local minimum analysis. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 2. AISC, vol. 84, pp. 353–359. Springer, Heidelberg (2010)
6. Mazurek, P.: Optimization of bayesian track-before-detect algorithms for GPGPUs implementations. Electrical Review R 86(7), 187–189 (2010)
7. Mazurek, P.: Suppression of impulse noise in track–before–detect algorithms. Computer Applications in Electrical Engineering 8, 201–211 (2010)
8. Mazurek, P.: Code reordering using local random extraction and insertion (LREI) operator for GPGPU-based track-before-detect systems. Soft Computing 18(6), 1095–1106 (2013)
9. Mazurek, P.: Track–before–detect filter banks for noise object tracking. International Journal of Electronics and Telecommunications 59(4), 325–330 (2013)
10. Mazurek, P.: Preprocessing using maximal autocovariance for spatio-temporal track-before-detect algorithm. In: Choras, R.S. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 45–54. Springer, Heidelberg (2014)
11. Stone, L., Barlow, C., Corwin, T.: Bayesian Multiple Target Tracking. Artech House (1999)

Hybrid Shape Descriptors for an Improved Weld Defect Retrieval in Radiographic Testing

Nafaa Nacereddine[1,2] and Djemel Ziou[3]

[1] Centre de Recherche en Soudage et Contrôle, Algiers, Algeria
[2] Lab. des Math. et leurs Interactions, Centre Univ. de Mila, Algeria
n.nacereddine@csc.dz
[3] Dpt. de Math. et Informatique, Univ. de Sherbrooke, Québec, Canada
djemel.ziou@usherbrooke.ca

Abstract. In this paper, four region-based shape descriptors well reported in the literature are used to characterize weld defect types of crack, lack of penetration, porosity and solid inclusion, usually encountered in radiographic testing of welds. The rectangularity and the roundness in the geometric descriptor (GEO) are used in order to propose an hybridization algorithm so that the hybrid descriptor issued from GEO and each of the other descriptors becomes more discriminant in such application where, due to bad radiographic image quality and weld defect typology, the human film interpretation is often inconsistent and labor intensive. According to the results given in the experiments, the efficiency of the proposed hybrid descriptors is confirmed on the weld defects mentioned above where, the retrieval scores are significantly improved compared to the original descriptors used separately.

1 Introduction

Objects are very often distinguishable on the basis of their visible features, such as shape, color, texture, etc., where, object's shape is recognized to play an important role in image retrieval or pattern recognition tasks. So, shape feature extraction consists in transforming an image in a description which is more easily usable. In general, the principal quality of a shape feature descriptor is its high discrimination ability, so that it can group similar shapes in the same class and separate the dissimilar shapes in different classes [1]. Moreover, ideal description must be stable, concise, unique, accessible and invariant to geometric transformations. There are several methods trying to quantify the shape as it is done by human intuition. Those can be classified broadly in two major families: statistical and structural methods, where the formers, which are the subject of this work, are based on the statistical study of measurements which we carry out on the shape to recognize [2]. With the statistical methods, each image is represented by an n-dimensional vector called descriptor. A review work, among many others, on the shape description techniques can be found in [3]. In weld radiographic testing (RT) [4], we can obtain radiograms in which the weld defects, if they exist, have various sizes and orientations because the same defect

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_16

can be viewed from several angles and this; according to the orientation and the distance of the irradiated welded component w.r.t. [1] the radiation source site. These practical considerations do not have to prevent, for example, that a crack should be identified as such whatever its size and its orientation [5]. This is why, the developed descriptors, to be effective, must at first be invariant to rotation, scaling and translation (RST). The purpose of this paper is then to use effective shape descriptors permitting to characterize the weld defects in the radiographic image so that they are recognized as elements of defect classes easily identifiable. Firstly, the region of interest (ROI), i.e. the radiogram region where the interpreter suspects the presence of flaws, is selected by user. Afterward, the ROI segmentation in defect and background regions is performed. The defect indication consists then in a unique region simply connected which should be described in terms of visual features. To this end, in one hand, four region-based shape descriptors are used, namely geometric descriptor (GEO)which will be developed later, generic Fourier descriptor (GFD)[6], Radon composite features (RCF)[7] and lastly, Zernike descriptor (ZD)[8]. The choice of these shape descriptors is motivated by their strength and well reported performances in the recognition of binary objects. They are mainly based on the geometric attributes, the transforms of Fourier and Radon, the polar transform and the Zernike moments. In the other hand, hybrid descriptors resulted from the combination of GEO with the others according to a strategy which will be unveiled later, are originally achieved to produce better outcomes for the weld defect image retrieval. The remainder of the paper is organized as follows. Sections 2 and 3 are respectively devoted to the segmentation and the binary shape description for weld defect radiographic images. The experiments are conducted in Section 4 where in addition, the algorithm implementing the hybrid descriptor is provided. Finally, Section 5 draws the concluding remarks.

2 Image Segmentation

In any computer vision system, the segmentation constitutes one of the most significant problems because the result, obtained at the end of this stage, strongly governs, depending on the application, the object classification rate, the image retrieval performance, the scene interpretation quality, etc. There is a wide variety of ways to achieve segmentation. Thousands of references related to segmentation can be found in literature describing hundreds of methods. Since there is no general solution with the problem of image segmentation, these techniques must often take in consideration the specificities of the image to treat, the type of visual indices to extract, the nature of the problem to be solved downstream of the segmentation and the exploiting constraints such algorithmic complexity, real time processing, material constraints related to acquisition system, storage capacity, etc. [9]. In our application, namely the weld radiographic image evaluation, the segmentation consists in the extraction of the weld defect, considered as

[1] w.r.t. means with respect to

an object on a background. Since the weld radiographic image is often characterized by, among others, bad contrast, non uniform background, blurry contours and high image noise, a region-based active contour represented by a cubic B-spline [10] is used in this work. The motivations of such a choice are detailed in the introduction of the mentioned reference. The binary objects representing the weld defects, on which a set of shape descriptors will be computed in the next section, are then easily drawn from the estimated contours.

3 Shape Descriptors for Weld Defect Images

In the shape analysis and recognition domains, it is of a particular interest to represent a binary object in $x - y$ plan by a function f as

$$f(x,y) = \begin{cases} 1 \text{ if } (x,y) \in D \\ 0 \text{ otherwise.} \end{cases} \tag{1}$$

where D is the binary object domain.

Other than GFD, RCF and ZM which are well defined in their corresponding papers cited in the introduction, we provide here the shape measures (some of them are defined in [11]) that compose GEO and which reveals to be invariant to RST transforms.

Compactness: It gives a numerical quantity representing the degree to which a shape is compact. It is measured as the ratio of the object area A to the area of a circle (the most compact shape) having the same perimeter P

$$C = \frac{4\pi A}{P^2} \tag{2}$$

Elongation: The minimum bounding box is the smallest rectangle that contains every point of the object. From it, the elongation E is defined as

$$E = 1 - \frac{W}{L} \tag{3}$$

where L and W are length and width of the minimum bounding box.

Rectangularity: It represents how rectangular a shape is, i.e. how much it fills its minimum bounding box. It is given by

$$R = \frac{A}{L \times W} \tag{4}$$

Ellipticity: It is noted ϵ and amounts to the ratio of the minor axis (b) to the major axis (a) of the object equivalent ellipse as follows

$$\epsilon = 1 - \frac{b}{a} \tag{5}$$

Cumulative signature: It is defined as in [11] by

$$S = \frac{1}{2\pi \max_{\theta}(r(\theta))} \int_0^{2\pi} r(\theta) d\theta \tag{6}$$

where $r(\theta)$ is the polar signature which is invariant to translation. The normalized integral version guarantees the invariance to rotation and scaling.

Roundness: The roundness ρ measures how closely the shape of an object approaches that of a circle. It is given by

$$\rho = \frac{4A}{\pi L_{max}^2} \tag{7}$$

where L_{max} is the longest chord with the chord is the line connecting two pixels on the object boundary.

4 Experiments

More than 340 weld defect regions are extracted from the radiographic films provided by the International Institute of Welding. These ROIs represent four weld defect classes: crack (CR), lack of penetration (LP), porosity (PO) and solid inclusion (SI). A sample of these defect types is illustrated in Fig. 1.

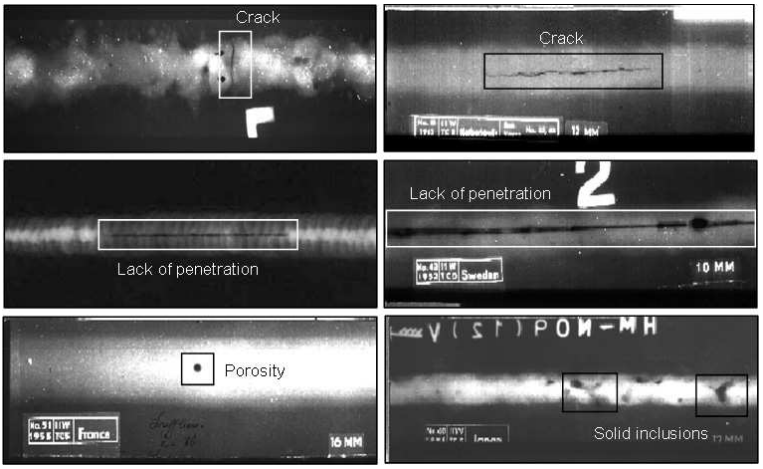


Fig. 1. A sample of ROIs representing CR, LP, PO and SI defects

After the segmentation step is achieved as shown in Fig. 2, the weld defect database is indexed by all the proposed descriptors, GEO, GFD, RCF and ZD. To perform any evaluation in the context of image retrieval, two major issues must be addressed: the acquisition of ground truth and the definition of performance criteria. To construct the ground truth, the overall weld defect image database is divided *a priori*, by a RT expert, in four image sets representing the weld defect classes mentioned above.

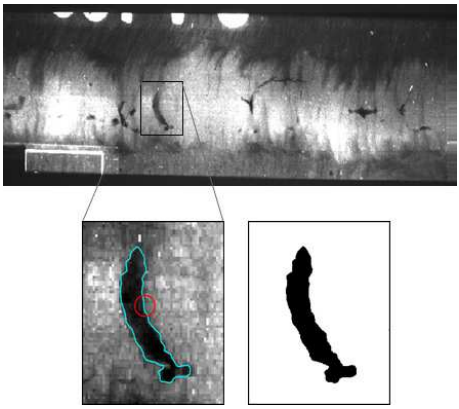


Fig. 2. Solid inclusion ROI and its segmentation results using the method in [10]. The initial and final contours are represented in red and cyan, respectively

For our experiments, to compare between descriptors in term of performance, each of the images in the database is used as a query to which all the images in the database are compared with. For any two weld defects I and J, the pattern-matching score is defined as the ℓ_1-norm distance

$$Dist(I, J) = \|D(I) - D(J)\|_1 \tag{8}$$

where $D \in \{\text{GEO}, \text{GFD}, \text{RCF}, \text{ZD}\}$.

The criterion used to compare the descriptor performances is the precision-recall curve [12]. At first, we compute in Fig. 3, for each descriptor, the confusion matrix to detect possible mislabeled defect classes.

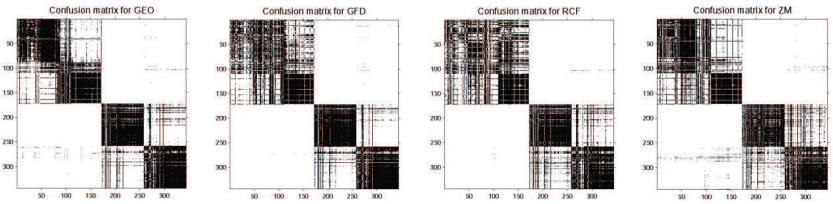


Fig. 3. Confusion matrix for each descriptor where the element sets #1-86, #87-172, #173-258 and #259-344 are, respectively, CR, LP, PO and SI indications

According to Fig. 3, except GEO descriptor, the problematic classes for the other descriptors are CR and LP which present a pronounced mismatching. Indeed, for GFD, RCF and ZD, the confusion matrix presents different class label squares in terms of size and degree of darkening about the two classes mentioned above. As shown in the recognition rates given in Fig. 5, the best discriminating descriptor for the crack class is GEO. In fact, CR and LP present

linear shape and it is difficult to discriminate between them. Nevertheless, some components of GEO, such as roundness and rectangularity, illustrated in Fig. 4 reveal to be decisive in the discriminating ability as it will be detailed in the following.

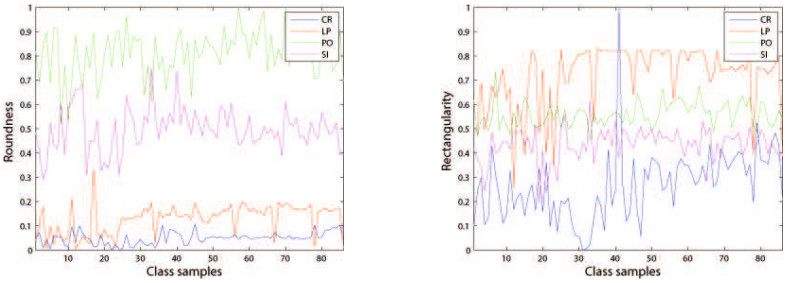


Fig. 4. Roundness and Rectangularity values for each weld defect class

With the aim to improve the weld defect retrieval results, we propose hybrid descriptors H(GEO,GFD), H(GEO,RCF), H(GEO,ZD) issued from the hybridization between GEO and GFD, RCF and ZM, respectively, according to a certain strategy. This latter consists in substituting each of GFD, RCF and ZD by n_d values of GEO (n_d is the cardinality of each class) having the lowest sum $\rho + R$ because, as obseved in Fig. 4, the crack, unlike the other defect types, is characterized by low compactness and low rectangularity and, accordingly, low sum of the both. Thus, the cracks will have the biggest chance to be represented by GEO values. We recall that to perform the hybridization operation summarized in Algorithm 5., GEO should be interpolated to reach the components number of each other descriptor.

The four first graphs in Fig. 5, illustrate the $pr - re$ curves for the descriptors and their hybrid versions for each weld defect class. The fifth graph draws the retrieval results averaged on the whole database for each of the simple or the hybrid descriptors. It is noticed from Fig. 5 that in general for the simple descriptor category, GEO outperforms all the others, except GFD for the porosity class. In fact, although the good behavior of GFD, RCF and ZD for the description of planar objects, the descriptor GEO is built so that each of its components has a relationship to the weld defect type. Explanations about these relationships are found in §2.1.1 of [11]. In the other hand, the hybrid descriptors outperform all the simple descriptors, except (1) GFD which is the best for the porosity class and (2) GEO which is comparable to them for the solid inclusion. Except, H(GEO,ZD) which is comparable to GEO for the whole weld defect database, the two others, i.e. H(GEO,GFD) and H(GEO,RCF) are the most performing for the application on which this paper is devoted, namely, weld defect image retrieval in radiographic testing.

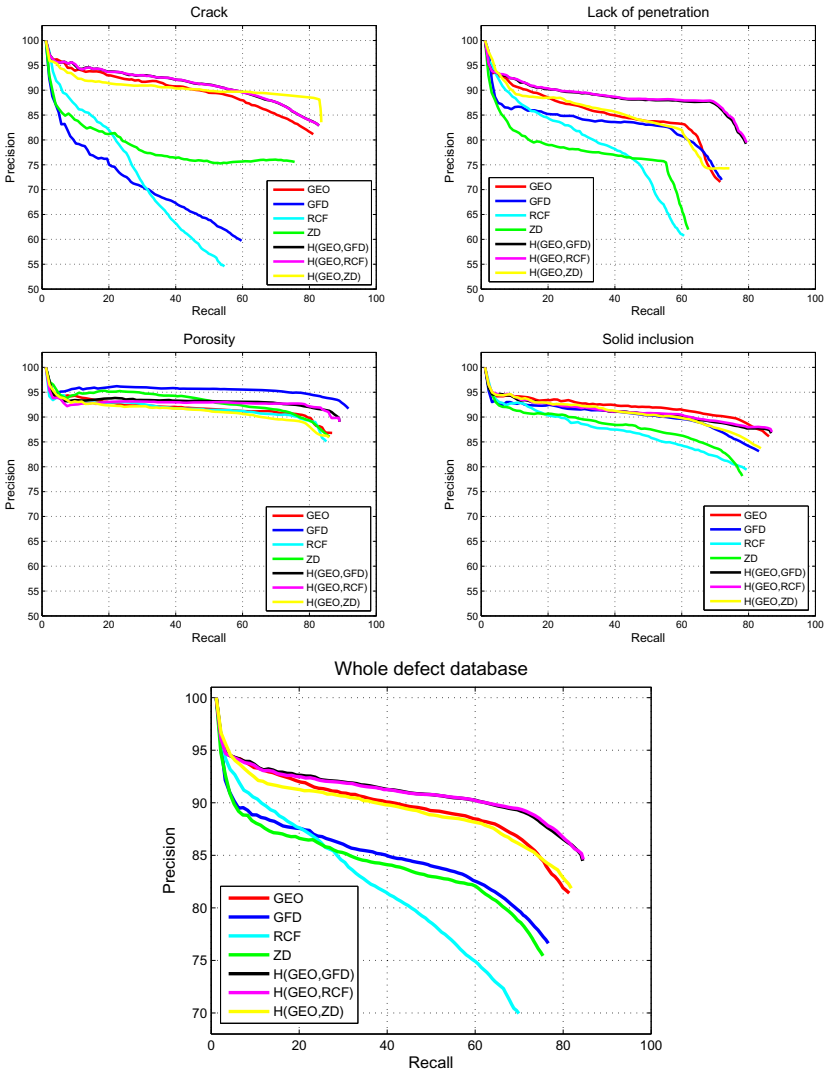


Fig. 5. Retrieval results for GEO, GFD, RCF and ZD and their hybrid versions

Algorithm 5. Computation of the Hybrid shape descriptor H(GEO,DSC)

Data: n_d : number of individuals (defects) by class
n_c : number of defect classes in database
DSC $\in$ {GFD,RCF,ZD}
$i, k = 1, \ldots, n_d$; $j = 1, \ldots, n_c$
size(GEO)=(n_d, l_{GEO}, n_c)
size(DSC)= (n_d, l_{DSC}, n_c)
$R(i,j) = \text{GEO}(i,3,j)$; $\rho(i,j) = \text{GEO}(i,6,j)$
Result: GEO_{int} by interpolation
H(GEO,DSC) by hybridization
Initialization: H(GEO,DSC) = DSC
$\text{GEO}_{INT} \Leftarrow$ GEO / size(GEO_{INT}) = size(DSC)
for *each* k **do**
  for *each* i **do**
    for *each* j **do**
      $(i_0, j_0) \leftarrow \arg\min_{i,j} (R(i,j) + \rho(i,j))$;
    end
  end
  H(GEO,DSC)$(i_0, :, j_0) = \text{GEO}_{int}(i_0, :, j_0)$;
  $R(i_0, j_0) + \rho(i_0, j_0) \leftarrow \eta/\eta \gg 1$;
end
N.A.: $n_d = 86$; $n_c = 4$; l_{GEO}=6; l_{DSC}=54,192,36 for GFD,RCF,ZD, resp.

5 Conclusion

With the aim to characterize the weld defect indications encountered in non-destructive tesing by radiography, in terms of shape, we have proposed in this paper, four descriptors, namely GEO, GFD, RCF and ZD, well known in the literature about their high discrimination ability for binary objects. It is shown from the retrieval results and the confusion matrices that GEO outperforms the rest of descriptors on the whole weld defect database and particularly on cracks. This interesting performance for GEO is due to the fact that its components are not chosen fortuitely but are related to the defect type. The rectagularity and the roundness measures are then exploited to build hybrid descriptors H(GEO,GFD), H(GEO,RCF) and H(GEO,ZD) formed from the comnination of GEO with each of the other simple descriptors. The outstanding performances of H(GEO,GFD), H(GEO,RCF) and to a lesser degree H(GEO,ZD), on the whole defect database, are indisputable as shown in the retrieval scores where, only the latter is comparable to GEO.

References

1. Kim, H., Kim, J.: Region-based shape descriptor invariant to rotation, scale and translation. Sig. Process.: Image Communication 16(1), 87–93 (2000)
2. Hastie, T., Tibshirani, T., Friedman, J.: The elements of statistical learning: Data Mining, Inference and Prediction. Springer Series in Statistics, NY, USA (2001)

3. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition 37(1), 1–19 (2004)
4. Richard, H., et al.: Nondestructive Testing Handbook, Moore, P.O (ed.), 3rd edn. Radiographic Testing, vol. 4. American Society of Nondestructive Testing ASNT, USA (2002)
5. Nacereddine, N., Tridi, M.: Computer-aided shape analysis and classification of weld defects in industrial radiography based invariant attributes and neural networks. In: Proc. of 4 th Int. Symp. on Image and Signal Processing and Analysis (ISPA 2005), Zagreb, Croatia, pp. 88–93 (2005)
6. Zhang, D., Lu, G.: Shape-based image retrieval using generic Fourier descriptor. Signal Processing: Image Communication 17(10), 825–848 (2002)
7. Chen, Y.W., Chen, Y.Q.: Invariant description and retrieval of planar shapes using Radon composite features. IEEE Trans. on Signal Processing 56(10), 4762–4771 (2008)
8. Khotanzad, A., Hong, Y.H.: Invariant image recognition by Zernike moments. IEEE Trans. on Patt. Anal. and Mach. Intell. 12(5), 489–497 (1990)
9. Soler, L., Malandrin, G., Delingette, H.: Segmentation automatique: Application aux angioscanners 3D. Revue de Traitement de Signal 15(5), 411–431 (1998)
10. Nacereddine, N., Hamami, L., Ziou, D., Goumeidane, A.B.: Adaptive B-spline model based probabilistic active contour for weld defect detection in radiographic imaging. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 2. AISC, vol. 84, pp. 289–297. Springer, Heidelberg (2010)
11. Nacereddine, N., Ziou, D., Hamami, L.: Fusion-based shape descriptor for weld defect radiographic image retrieval. International Journal of Advanced Manufacturing Technology 68(9-12), 2815–2832 (2013)
12. Smith, J.R.: Image retrieval evaluation. In: Proc. of IEEE Workshop on Content-based Access of Image and Video Libraries (1998)

On the Usefulness of Combined Metrics for 3D Image Quality Assessment

Krzysztof Okarma

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

Abstract. Due to the growing popularity of 3D imaging technology which can be observed in recent years, one of the most relevant challenges for image quality assessment methods has become their extension towards reliable evaluation of stereoscopic images. Since known 2D image quality metrics are not necessarily well correlated with subjective quality scores of 3D images and the exact mechanism of the 3D quality perception is still unknown, there is a need of developing some new metrics better correlated with subjective perception of various distortions in 3D images. Since a promising direction of such research is related with the application of the combined metrics, the possibilities of their optimization for the 3D images are discussed in this paper together with experimental results obtained for the recently developed LIVE 3D Image Quality Database.

1 Motivation

Analysis of stereoscopic 3D images is currently a very active area of research which has developed a number of algorithms and working systems. A natural consequence of the development of stereovision algorithms is extremely important need to develop methods to assess the quality of stereoscopic 3D images e.g. due to some limitations of communication is stereovision based systems. However, this task is not a straightforward extension of the methods used to assess the quality of 2D images, because apart from the quality assessment of each image of the pair, it is necessary to assess the quality in terms of the perception of 3D scenes. As in the case of a reliable quality assessment of color images [1], as well as video sequences [2], it is still an open field of research with a growing number of publications in leading journals and scientific conferences recognized worldwide by researchers interested in the field of image processing and analysis.

Research activities on the quality of stereoscopic images are conducted in several centers in the world although they are mostly related to image quality for cinema and television purposes where cinema production a complex process of obtaining and processing stereoscopic images is used. In addition, the captured image is processed in computer post-production in order to highlight the

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_17

3D effects in a certain part of the image. In the case of television production, even when its real-time implementation, it can be treated as a fully controlled process without any unpredictable components with limited possibility of distortions. Nevertheless, in many applications e.g. related to telerobotics, some image distortions may appear and their character may be completely different than it takes place in the movies or 3DTV.

Until recent years, most of the known image quality metrics was limited to 2D images in grayscale, which was the effect of availability of test databases necessary to verify the proposed objective metrics in compliance with subjective scores. In the past few years there has been a rapid development of quality assessment methods for images and video sequences partly due to the development of new databases of test images and video sequences, containing also the subjective scores. Among these databases one can also find some images containing several types of distortions occurring at the same time such as LIVE Multiply Distorted Image Quality Database [3], and video sequence databases developed by the Laboratory for Image and Video Engineering (LIVE), EPFL/PoliMi, IRCCyN/IVC or MMSP.

Nevertheless, the most important from the point of view of the development of methods for assessing the quality of stereoscopic images are the databases of 3D images and 3D video sequences. The first of this kind of database, published in 2008, is IRCCyN/IVC 3D Image Quality Database [4] containing blurry images as well as JPEG and JPEG2000 compressed ones. However, in subsequent years, some other databases have appeared containing a larger number of distortions, such as MMSP 3D Image Quality Assessment Database and MMSP 3D Video Quality Assessment Database for different interocular distances between the cameras, IRCCyN/IVC DIBR Image Quality Database for different depth-image based rendering algorithms. In addition, over the past two years three new datasets have been made available: IRCCyN/IVC NAMA3DS - COSPAD1 3D Video Quality Database, IVP Anaglyph Image Database and the most relevant LIVE 3D Image Quality Database [5,6] used in this paper. In addition, there are also some other databases available but they do not contain any results of subjective evaluations, such as e.g. RMIT 3D Video Library, which are useless for the development of 3D image quality metrics.

2 Recent Attempts to 3D Image Quality Assessment

Many new ideas related to image and video quality assessment, also for 3D images, are presented each year at the International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM) usually held in Scottsdale, Arizona. Analyzing the published proceedings from recent years, growing interest of 3D image quality assessment can be easily noticed. An example of such metrics, proposed by You et al. [7], is based on the integration of disparity information and 2D metrics by calculation of two quality maps which are further combined locally. The mean is taken as an intermediate quality of the distorted image and finally is combined with the quality of the disparity image.

Some other state-of-the-art 3D image quality assessment metrics have also been proposed by various researchers during last five years. An early example is the metric proposed by Gorley and Holliman [8] based on asymmetric compression of stereoscopic signals and asymmetric distortions. Benoit et al. also proposed the fusion of 2D quality metrics and the depth information [4] using the combination of well-known Structural Similarity (SSIM) [9] and C4 metric exploiting an implementation of an elaborated model of the human visual system. One of their (still up-to-date) conclusions is that "the depth information can improve the quality metric but the relation with image 'naturalness', 'viewing experience' and 'presence' has still to be investigated in depth, depending also in the different 3D display technology used".

Considering the necessity of changing parameters of the video transmission "on-the-fly", the reduced-reference metric for 3D depth map transmission using the extracted edge information has been proposed by Hewage et al. [10]. The authors of this paper noticed that edges and contours of the depth map can represent different depth levels and hence can be used in quality evaluations. Nevertheless, since this metric is based on the PSNR, assuming the binary character of edge information, it may be considered as a good starting point for further extensions. Similarly, the metric proposed by Yang et al. [11] is also based on the PSNR which is supplied by the Stereo Sense Assessment (SSA) but the main advantage of this method is the possibility of fast calculation. An alternative method proposed by Zhu and Wang [12] uses a multi-channel vision model based on 3D wavelet decomposition which is more perceptually consistent than e.g. PSNR or MSE. The main idea is based on the wavelet-based perceptual decomposition of the video sequences into three domains (temporal, spatial and view domain) and further contrast conversion, masking and nonlinear summation. Another idea has been proposed by Shen et al. [13] for estimation of stereo image quality based on a multiple channel Human Visual System (HVS) for use in image compression. Proposed metric utilizes the contourlet transform but the results presented by the authors are quite preliminary.

Considering the demands of objective no-reference ("blind") 3D image quality assessment metrics which do not require the knowledge of an original undistorted image, the first attempts in this field have also been made. An example metric based on the assumption that the perceived distortion and disparity of any stereoscopic display is strongly dependent on local features, such as edge (non-plane) and non-edge (plane) areas, has been proposed by Akhter et al. [14]. The metric has been developed for JPEG coded stereoscopic images and is based on segmented local features of artifacts and disparity.

3 Application of Combined Metrics

Due to the presence of various approaches to the problem of 3D image quality assessment, currently there is no single universal metric which could be applied for most 3D images, even containing only some most typical distortions. Since many of the existing solutions mentioned above use quite complicated models

utilizing both the image information and the reference depth, they may not be attractive for many technical systems e.g. assuming real-time video transmission. For those reasons the goal if this paper is related to the verification of the usefulness of relatively simple approach based on multi-method fusion leading to combined metrics, previously successfully applied for the 2D images.

The idea of the combined metrics is based on the nonlinear combination of three or more image quality assessment methods of different kinds. The main motivation of such an approach is the necessity of nonlinear mapping of the raw quality scores which is conducted by many researchers, typically using the logistic function. Avoiding such mapping is possible by including the nonlinearity inside the computational procedure by weighted product of various metrics leading to combined (hybrid) metrics.

The first such approach denoted as Combined Quality Metric (CQM) [15] has been proposed as the combination of three metrics: Multi-Scale Structural Similarity [16], Visual Information Fidelity [17] and R-SVD metric [18] with weights optimized using the largest available image quality assessment database known as Tampere Image Database (TID2008). Further replacement of the R-SVD metric by Feature Similarity (FSIM) [19] has led to the Combined Image Similarity Index [20] achieving high correlation with subjective scores for the most relevant datasets. Similar approach has also been presented in the paper [21] where a deeper analysis of many combinations has been conducted. Nevertheless, the proposed multi-metric fusion is based on the machine learning approach decreasing the universality of this particular solution. Another modification of the combined metric (Extended Hybrid Image Similarity - EHIS) has been analyzed in the paper [22] which is based on the modifications of the FSIM metric together with Riesz-based Feature Similarity [23] proposed earlier by the the FSIM inventors.

Since the main goal of the paper is related to the verification of usefulness of combined metrics, the experiments have been focused on the calculation of correlation coefficients with subjective evaluations for the metrics utilizing image data without the use of depth information which are available only for reference images. Such assumption makes it possible to compare the consistency of all metrics with subjective perception of image distortions as the depth information remains unchanged for the images included in LIVE 3D Image Quality Database used for the experiments. The dataset has been chosen due to rather poor definitions of stereoscopic distortions in all currently available databases and the significant differences between them.

One of the main advantages of the LIVE 3D database is its division into two independent sets (Phase I and Phase II) containing symmetrical and asymmetrical distortions respectively. Due to such construction of the database the combined metrics can be independently analyzed for those two kinds of distortions since much better correlation of existing 2D metrics with provided Differential Mean Opinion Scores (DMOS) should be expected for the 'symmetrical' set (Phase I).

4 Details of Experiments and Results

In the conducted experiments some well-known 2D image quality metrics have been used together with the combined ones such as mentioned CQM [15], CISI [20] and EHIS [22]. Since the color image quality assessment is a separate problem, not discussed in this paper, all the metrics (except FSIMc being the color version of Feature Similarity) have been applied for the images converted to grayscale according to the widely used ITU BT.601-7 recommendation:

$$Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \ , \tag{1}$$

applied e.g. in MATLAB's *rgb2gray* function.

In the first part of experiments, conducted independently for both parts of the database, the Pearson's Linear Correlation Coefficients (PCC) with DMOS values have been calculated indicating the quality prediction accuracy (for the average raw scores obtained for the left and right images without any nonlinear mapping), as well as Spearman Rank Order Correlation Coefficients (SROCC) for verification of metrics' prediction monotonicity. The results obtained for some well-known metrics and three combined ones assuming the default weights proposed in respective papers (typically obtained as the result of optimization for the well-known TID2008 database) are presented in Table 1. Conducting the optimization of the weighting exponents for all combined metrics further increase of the PCC and SROCC values can be gained and the obtained results are shown in Table 2.

Analyzing the values of the correlation coefficients presented in Table 1 it can be noticed that the direct application of the combined metrics for the 'symmetrical' subset (Phase I) does not lead to satisfactory results, especially for the EHIS metric. Nevertheless, for the asymmetrically distorted Phase II images the best results have been achieved for the CISI metric with default weights [20] both by means of prediction accuracy and monotonicity.

Considering the different perception of distortions in the 2D and 3D images, the optimization of weights have been conducted, using MATLAB's *fminsearch* and *fminunc* functions, for the combined metrics and meaningful increase of the correlation with subjective scores has been achieved for all of them as shown in Table 2. Nevertheless, it is worth noticing that the optimization of weights using the Phase II subset leads to slightly worse results for the 'symmetrical' images and reversely. The smallest differences can be noticed for the CQM metric.

For the Phase I images the increase of the PCC values to over 0.93 can be obtained for each of combined metrics with only 0.915 for the FSIM whereas the SROCC values remain similar. For the Phase II images, characterized by different perception of distortions due to their asymmetry, the increase of the Pearson's correlation to 0.8953 can be observed for the EHIS metric with 0.8464 for the Gradient SSIM. Spearman correlation increases as well, from 0.833 to over 0.89 respectively.

Table 1. Obtained results of the PCC and SROCC values for various single metrics and the combined metrics with default weights

Subset	Phase I		Phase II	
Metric	PCC	SROCC	PCC	SROCC
SSIM	0.86127	0.87724	0.78463	0.79320
MS-SSIM	**0.87387**	0.89747	0.72748	0.73131
GSSIM	**0.89366**	0.91174	**0.84639**	**0.83305**
QILV	0.64446	0.91385	0.42955	0.67806
IW-SSIM	**0.89300**	**0.93353**	0.63632	0.74955
R-SVD	0.54962	0.52507	0.63548	0.61600
IFC	0.46750	0.91981	0.52732	0.76531
VIF	**0.86985**	**0.92054**	0.83407	0.81641
VIFp	0.84215	0.91357	0.80501	0.80557
VSNR	0.77026	0.88202	0.66007	0.73344
RFSIM	0.75834	0.82678	0.79863	0.78687
WFSIM	**0.89228**	**0.93033**	0.68225	0.74970
FSIM	**0.91504**	**0.92769**	0.76984	0.78638
FSIMc	**0.91389**	**0.92698**	0.78245	0.79771
CQM	0.74246	**0.92079**	0.46043	0.74973
CISI	**0.86599**	**0.92718**	**0.86137**	**0.84771**
EHIS	0.13234	0.43470	0.39562	0.52314

Table 2. Obtained results of the PCC and SROCC values for the optimized combined metrics

Subset	Phase I		Phase II	
Metric	PCC	SROCC	PCC	SROCC
CQM_{opt1}	**0.93358**	0.92813	**0.85183**	**0.86375**
CQM_{opt2}	0.88153	0.91301	**0.88968**	**0.88251**
$CISI_{opt1}$	**0.93444**	**0.93000**	0.81918	0.81587
$CISI_{opt2}$	0.88281	0.90219	**0.87239**	**0.86717**
$EHIS_{opt1}$	**0.93618**	**0.93286**	0.82138	0.81956
$EHIS_{opt2}$	0.89774	0.91867	**0.89528**	**0.89090**

5 Concluding Remarks

Application of the combined metrics for 3D image quality assessment leads to meaningful increase of quality prediction accuracy and monotonicity, expressed as the Pearson's and Spearman's correlation coefficients respectively. The results provided in this paper confirm both the usefulness of the hybrid metrics also for 3D images and the different perception of similar distortions present in 2D and 3D images, especially applied in an asymmetric way for the left and right images from the consecutive stereoscopic pairs.

It is worth to notice that the direct use of well-known metrics may lead to relatively good performance for symmetrically distorted images whereas the presence of asymmetric distortions causes a significant decrease of the PCC and SROCC values for most single metrics. Nevertheless, optimization of the weighting coefficients for the combined metrics allows to obtain much better performance than for any of single metrics used in the experiments.

Since in the conducted experiments no information related to depth has been utilized, a natural direction of further research is the combination of the developed metrics with depth or disparity information provided in most of available databases, including the LIVE 3D Image Database used in the paper.

References

1. Okarma, K.: Colour image quality assessment using Structural Similarity index and Singular Value Decomposition. In: Bolc, L., Kulikowski, J.L., Wojciechowski, K. (eds.) ICCVG 2008. LNCS, vol. 5337, pp. 55–65. Springer, Heidelberg (2009)
2. 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)
3. Jayaraman, D., Mittal, A., Moorthy, A.K., Bovik, A.C.: Objective image quality assessment of multiply distorted images. In: Conf. Rec. 46th Asilomar Conf. Signals, Systems and Computers, pp. 1693–1697 (2012)
4. Benoit, A., LeCallet, P., Campisi, P., Cousseau, R.: Quality assessment of stereoscopic images. EURASIP Journal on Image and Video Processing Article ID 659024, 13 (2008)
5. Chen, M.-J., Su, C.-C., Kwon, D.-K., Cormack, L.K., Bovik, A.C.: Full-reference quality assessment of stereopairs accounting for rivalry. Signal Processing: Image Communication 28(9), 1143–1155 (2013)
6. Chen, M.-J., Cormack, L.K., Bovik, A.C.: No-reference quality assessment of natural stereopairs. IEEE Trans. Image Process. 22(9), 3379–3391 (2013)
7. You, J., Xing, L., Perkis, A., Wang, X.: Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. In: Proc. 5th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), pp. 61–66 (2010)
8. Gorley, P., Holliman, N.: Stereoscopic image quality metrics and compression. In: Proc. SPIE, vol. 6803, p. 5 (2008)
9. 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)
10. Hewage, C.: Reduced-reference quality metric for 3D depth map transmission. In: Proc. 4th 3DTV Conf.: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1–4 (2010)
11. Yang, J., Hou, C., Zhou, Y., Zhang, Z., Guo, J.: Objective quality assessment method of stereo images. In: Proc. 3rd 3DTV Conf.: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1–4 (2009)
12. Zhu, Z., Wang, Y.: Perceptual distortion metric for stereo video quality evaluation. WSEAS Trans. Signal Process. 5(7), 241–250 (2009)

13. Shen, L., Yang, J., Zhang, Z.: Stereo picture quality estimation based on a multiple channel HVS model. In: Proc. 2nd IEEE Int. Congress on Image and Signal Processing (CISP), pp. 1–4 (2009)
14. Akhter, R., Parvez Sazzad, Z., Horita, Y., Baltes, J.: No-Reference stereoscopic image quality assessment. In: Proc. SPIE. Stereoscopic Displays and Applications XXI, vol. 7524, p. 75240T (2010)
15. 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, Part I. LNCS (LNAI), vol. 6113, pp. 539–546. Springer, Heidelberg (2010)
16. Wang, Z., Simoncelli, E., Bovik, A.C.: Multi-Scale Structural Similarity for image quality assessment. In: Proc. 37th IEEE Asilomar Conf. on Signals, Systems and Computers, pp. 1398–1402 (2003)
17. Sheikh, H., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. 15(2), 430–444 (2006)
18. Mansouri, A., Mahmoudi-Aznaveh, A., Torkamani-Azar, F., Jahanshahi, J.: Image quality assessment using the Singular Value Decomposition theorem. Opt. Rev. 16(2), 49–53 (2009)
19. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A Feature Similarity index for image quality assessment. IEEE Trans. Image Proc. 20(8), 2378–2386 (2011)
20. Okarma, K.: Combined image similarity index. Opt. Rev. 19(5), 249–254 (2012)
21. Liu, T.-J., Lin, W., Kuo, C.-C.J.: Image quality assessment using multi-method fusion. IEEE Trans. Image Process. 22(5), 1793–1807 (2013)
22. Okarma, K.: Extended Hybrid Image Similarity - combined full-reference image quality metric linearly correlated with subjective scores. Elektronika ir Elektrotechnika 19(10), 129–132 (2013)
23. Zhang, L., Zhang, L., Mou, X.: RFSIM: A feature based image quality assessment metric using Riesz transforms. In: Proc. 17th IEEE Int. Conf. Image Processing, pp. 321–324 (2010)
24. International Telecommunication Union: Recommendation BT.601-7 - Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios (2011)

Application of Super-Resolution Algorithms for the Navigation of Autonomous Mobile Robots

Krzysztof Okarma, Mateusz Tecław, and Piotr Lech

West Pomeranian University of Technology, Szczecin
Faculty of Electrical Engineering
Department of Signal Processing and Multimedia Engineering
26. Kwietnia 10, 71-126 Szczecin, Poland
{krzysztof.okarma,mateusz.teclaw,piotr.lech}@zut.edu.pl

Abstract. In the paper the idea of using the super-resolution algorithms for the self-localization and vision based navigation of autonomous mobile robots is discussed. Since such task is often limited both by the limited resolution of the mounted video camera as well as the available computational resources, a typical approach for video based navigation of mobile robots, similarly as many small flying robots (drones), is using low resolution cameras equipped with average class lenses. The images captured by such video system should be further processed in order to extract the data useful for real-time control of robot's motion. In some simplified systems such navigation, especially in the within an enclosed environment (interior), is based on the edge and corner detection and binary image analysis, which could be troublesome for low resolution images.

Considering the possibilities of obtaining higher resolution images from low resolution image sequences, the accuracy of such edge and corner detections may be improved by the application of super-resolution algorithms. In order to verify the usefulness of such approach some experiments have been conducted based on the processing of the captured sequences of the HD images further downsampled and reconstructed using the super-resolution algorithms. Obtained results have been reported in the last section of the paper.

1 Introduction

Vision based autonomous mobile robot navigation becomes more and more interesting field of research during recent years. Growing interest in using real-time image and video analysis methods for the control of autonomous guided vehicles (AGVs) or unmanned aerial vehicles (UAVs) as well as many other robotic applications is caused mainly by the increase of the available computational power of processors and availability of relatively cheap cameras. Hence computer vision methods, together with cameras treated as complex passive sensors, can be considered as very attractive supplement to active or passive solutions typically used in robotics, such as infrared, laser or sonar sensors [1].

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_18

The advantages of vision systems in comparison to active sensors which alter the environment are well known, similarly as some limitations of GPS solutions, laser range finders or sonars (e.g. poor angular resolution). Therefore some applications of vision for the navigation of mobile robots have been considered even in 1980s and 1990s. Due to the development of cameras and video processing systems in consecutive years, some classical algorithms, often based on binary image analysis, have become more complex. Nevertheless, many solutions and algorithms are dedicated to a limited area of applications e.g. indoor navigation in corridors [2] or underwater navigation [3,4]. Some more advanced concepts may be based e.g. on saliency and visual attention [5] or Scale Invariant Feature Transform (SIFT) [6]. A comprehensive survey of many approaches can also be found in the paper [7].

Considering the types of navigation task (indoor or outdoor environment, map based, mapless or map building approach, structured or unstructured environments), the video analysis in low cost mobile robots is often conducted using simplified representation of images. Hence, the accuracy of edge and corner detection as well as their assignment in consecutive video frames is still a crucial element both in indoor and outdoor navigation.

For low cost autonomous robots (e.g. aerial ones), equipped with light low resolution cameras, high accuracy of self-positioning and navigation may be hard to achieve due to the physical limitations of the computer vision system. Nevertheless, especially in systems where a slight delay (less than a second) is acceptable, the accuracy of detection of some landmarks, edges or corners may be improved by using the super-resolution (SR) methods allowing the reconstruction of a single high resolution image based on a number of low resolution video frames assuming the presence of subpixel displacements in consecutive frames. Such situation is quite typical for a moving robot so the application of the super-resolution algorithms seems to be an interesting idea which can be useful for some of the mobile robot navigation systems.

2 Super-Resolution Algorithms

A typical approach to enlarging the digital images is based on the interpolation, using different methods e.g. nearest-neighbor, Lanczos resampling, bilinear or bicubic interpolation, supersampling or more advanced algorithms, such as patented S-Spline. Those algorithms are mostly based only on a single image in contrast to SR algorithms utilizing several consecutive video frames utilizing the sub-pixel shifts between multiple low resolution images representing the same scene, leading to the increase of the physical resolution of the output image.

During last several years numerous super-resolution algorithms have been proposed, including frequency domain algorithms and typically more computationally complex spatial domain methods which are less sensitive to model errors. Since a typical SR algorithm consists of two main steps related to image registration with motion estimation and image reconstruction, usually the latter is different in many algorithms. Image registration may be based on the analysis of the amplitude and phase frequency characteristics [8,9] or expansion into

Taylor series [10]. Some of the most popular approaches to image reconstruction algorithms are: iterated back-projection proposed by Irani and Peleg [11,12], Projection Onto Convex Sets (POCS), Papoulis-Gerchberg method [13], robust super-resolution [14] and normalized convolution [15].

Nevertheless, due to the computational demands of most of those methods, the bicubic interpolation combined with the image registration algorithm proposed in the paper [8] has been chosen. Considering also the possibilities of effective implementation of the algorithm in autonomous mobile robots, verification of the usefulness of the SR algorithms for the preprocessing of images used for mobile robot navigation purposes has been limited to algorithms with relatively low computational and memory demands.

3 Experimental Verification

The experiments have been conducted using two Full HD video sequences captured by the camera mounted on the Mobot-Explorer-A1 mobile robot. The first one has been captured in the corridor and the second one is an outside one. Such obtained frames have been considered as the reference and the video sequences have been downsampled by 3 in order to obtain low resolution frames being the input for the SR algorithm. For the comparison with the application of typical bicubic interpolation, the Structural Similarity (SSIM) metric [16] values have been calculated for the interpolated and the SR reconstructed images using the original Full HD frames as the reference.

Considering the fact that the most relevant areas of images for robot navigation purposes are located near the edges present in the consecutive video frames, a modification of the SSIM metric known as Three-Component SSIM (3-SSIM) [17] has been adapted for the additional verification of the impact of the SR algorithm on the obtained results. Calculation of the 3-SSIM metric is conducted separately for the three types of regions representing edges, textures and flat image areas. Since the most important regions in view of robot navigation are located near the edges the images have been masked by the result of morphological dilation of the edge detection result obtained using well known Canny filter. Then, the local values of the SSIM metric have been calculated only for those regions and finally averaged in order to obtain the final similarity score. The exemplary frames from both video sequences together with results of dilated edge filtering (shown as negatives) are presented in Fig. 1. As can be observed, the indoor frame is much less complex and contains smaller amount of edge information so the navigation task should be much simpler.

Since one of the relevant parameters of each super-resolution method is the number of input frames, all the experiments have been conducted for 3, 5, 7 and 10 frames used for the Full HD image reconstruction. Another important issue is related to the relative position of the reconstructed frame as better results are obtained for the middle frames from the sequence. Nevertheless, in order to reduce the amount of presented data, only the average values of the similarity scores calculated for each reconstructed frame are presented in Fig. 2. Almost

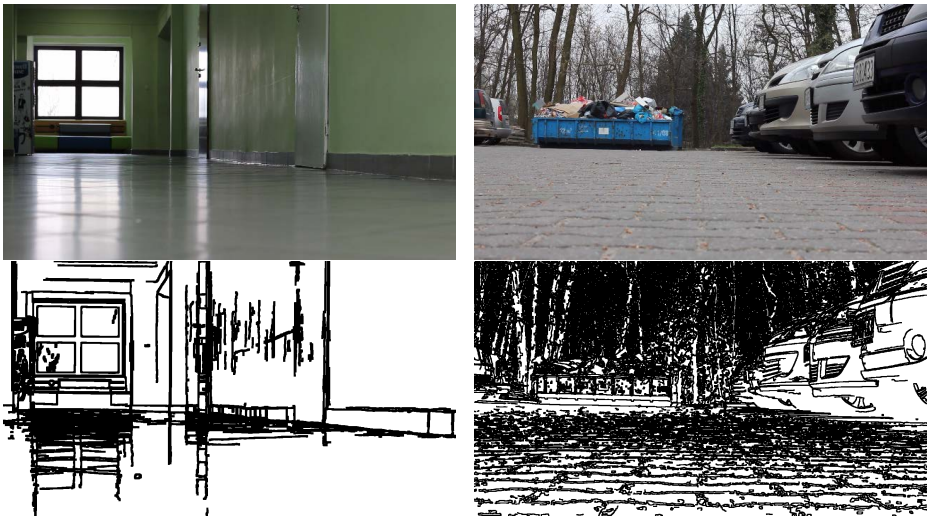


Fig. 1. Exemplary video frames for the indoor and outdoor sequences with results of Canny edge detection after dilation using 5×5 pixels structuring element

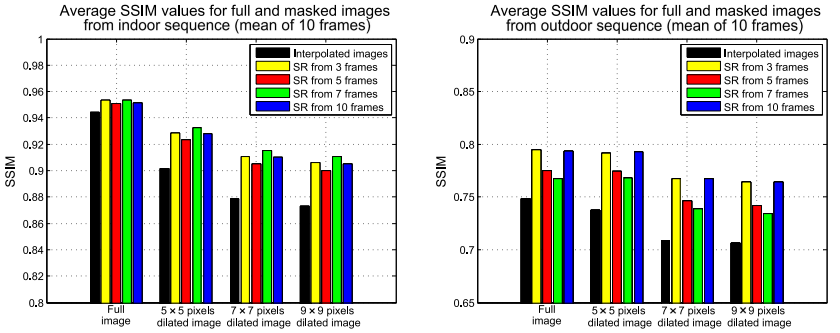


Fig. 2. Illustration of average SSIM and adapted 3-SSIM values calculated for full images and masked by dilated edges detected using Canny filter

identical results obtained for masking by 7×7 and 9×9 pixels structuring element for the morphological dilation of edges for the outdoor sequence are caused by the presence of many details on the image. Therefore almost whole area of the image has been covered by the obtained mask in both cases. Exemplary images obtained using the bicubic interpolation and SR algorithm are shown in Fig. 3 (for better visualization of details only magnified fragments of images are presented).

As the preprocessed images captured by the camera typically used for mobile robot navigation purposes are further subjected to edge detection in order to detect some specific lines, corners or landmarks on the obtained binary image, the additional verification has been conducted by comparison of such processed

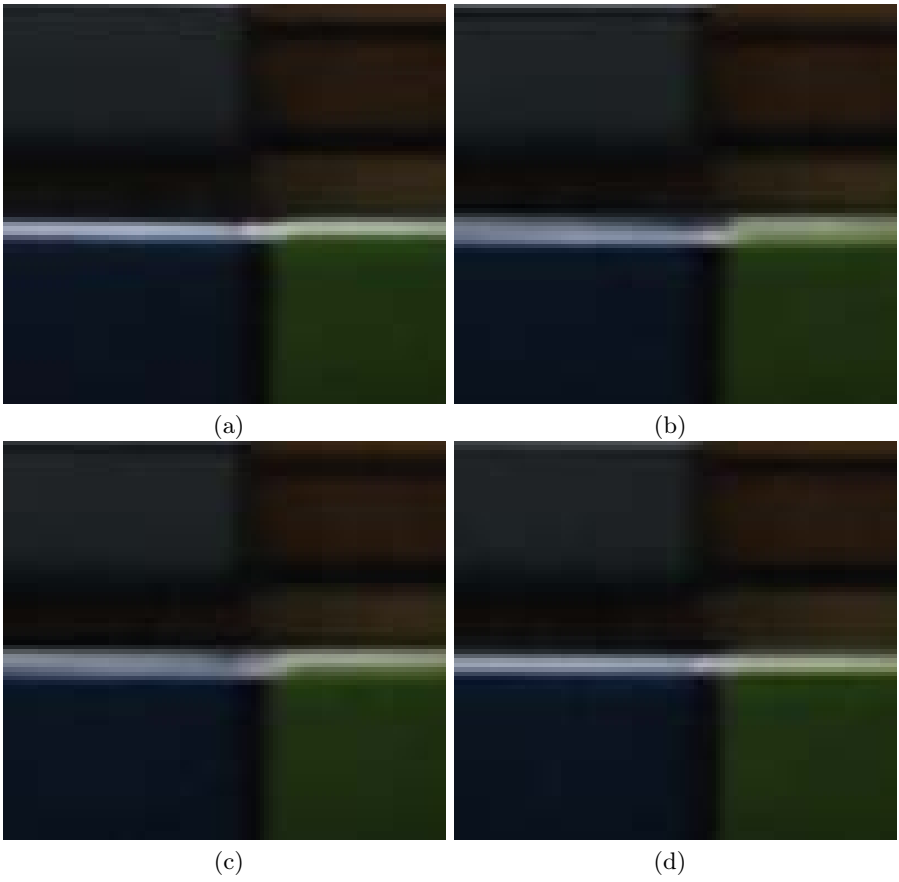


Fig. 3. Magnified fragments of an exemplary reference image from the indoor sequence (a) and exemplary images obtained using the bicubic interpolation (b) and SR algorithm using 3 and 10 frames – (c) and (d) respectively

images, regardless of the results illustrated in Fig. 3d where the advantages of the SR algorithm are clearly visible. For this purpose the reference frames and all reconstructed images have been filtered using Canny edge detecting filter and then the comparison of output images obtained for the reconstructed frames with the result obtained for the reference image has been done. In fact, the influence of the usage of the SR algorithms on the obtained binary images may be relevant not only for the navigation of autonomous robots but also for many machine vision embedded applications, often using specialized binary image codecs [18].

Since the comparison of the binary images using the correlation coefficients or image quality (or similarity) metrics such as e.g. Structural Similarity [16] may be inadequate, some metrics typically used in binary classification have been applied. Similar metrics are typically used also for the verification of image binarization algorithms. Comparing the binary images woth the reference one,

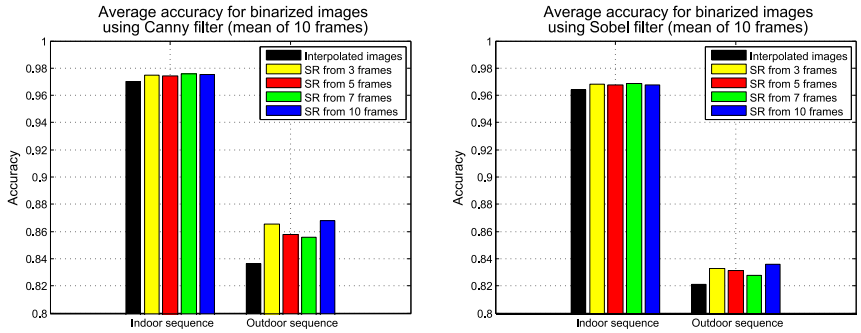


Fig. 4. Illustration of average edge detection accuracy in comparison to the reference image for reconstructed images subjected to Canny and Sobel filters

it is easy to determine the number of true positives (TP), true negatives (TN), false positives (FN) and false negatives (FN), leading to precision and recall as well as the F-Measure, specificity and accuracy and some other metrics such as e.g. Matthews correlation coefficient (MCC) [19]. Since all the metrics lead to similar conclusions, we have focused on the accuracy defined as:

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN} \,. \tag{1}$$

The experimental results obtained using Canny filter are illustrated in Fig. 4 as the average edge detection accuracy for the indoor and outdoor sequences achieved for the interpolated images and the outputs of the SR algorithm based on different numbers of frames. Higher values of the edge detection accuracy can be considered as the confirmation of the increase of the self-positioning (according to the specified patterns visible on the image) and vision based navigation accuracy of the mobile robot.

4 Conclusions and Future Work

Analyzing the results obtained for two acquired "raw" video sequences without any preprocessing operations, a strong influence of details present on the image can be easily noticed. The sequences chosen for the experiments can be considered as "hard" for the robot navigation purposes mainly due to the light conditions causing reflections of light on the floor and the presence of and many details (sett paved road, trees). Nevertheless, achieved results are promising both by means of the SSIM metric's values for the most relevant fragments of images reconstructed using the super-resolution algorithm, presented in Fig. 2, and the edge detection accuracy. The advantages of the SR algorithm can also be observed in Fig. 3 where the representation of the seat's horizontal edge is much more blurred for the bicubic interpolation (Fig. 3b) and the image obtained using the SR algorithm based only on three frames (Fig. 3c).

Due to the presence of many details in the outdoor sequence, the results of edge detection for the reconstructed frames are much worse than obtained for the indoor movie. Nevertheless, analyzing the results shown in Fig. 4 a greater relative improvement of the edge detection accuracy can be observed applying Canny filter for the outdoor video frames. Slightly worse results can be observed for Sobel filter with binary output.

Considering the assumed choice of the SR algorithm in view of its computational demand, caused by the specific area of applications, our future experiments will concentrate on the selection and possible modifications of some other image registration and reconstruction methods leading to further increase of the obtained accuracy. Further implementation of the chosen algorithms in the robotic system will allow the experimental verification of the dynamic properties of chosen methods for the real-time robot navigation.

References

1. Chatterjee, A., Rakshit, A., Singh, N.N.: Vision Based Autonomous Robot Navigation. SCI, vol. 455. Springer, Heidelberg (2013)
2. DeSouza, G.N., Kak, A.C.: Vision for mobile robot navigation: A survey. IEEE Trans. Pattern Anal. Machine Intell. 24(2), 237–267 (2002)
3. Dalgleish, F.R., Tetlow, S.W., Allwood, R.L.: Vision-based navigation of unmanned underwater vehicles: A survey. Part I: vision based cable-, pipeline- and fish tracking. Proc. Inst. Marine Engineering, Science and Technology. Part B, Journal of Marine Design and Operations B(7), 51–56 (2004)
4. Dalgleish, F.R., Tetlow, S.W., Allwood, R.L.: Vision-based navigation of unmanned underwater vehicles: A survey. Part II: vision based station keeping and positioning. Proc. Inst. Marine Engineering, Science and Technology. Part B, Journal of Marine Design and Operations B(8), 13–19 (2004)
5. Chang, C.-K., Siagian, C., Itti, L.: Mobile robot vision navigation & localization using Gist and Saliency. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), pp. 4147–4154 (2010)
6. Se, S., Lowe, D., Little, J.: Vision-based global localization and mapping for mobile robots. IEEE Trans. Robotics 21(3), 364–375 (2005)
7. Bonon-Font, F., Ortiz, A., Oliver, G.: Visual navigation for mobile robots: A survey. Journal of Intelligent and Robotic Systems 53(3), 263–296 (2008)
8. Vandewalle, P., Süsstrunk, S., Vetterli, M.: A frequency domain approach to registration of aliased images with application to super-resolution. EURASIP Journal on Applied Signal Processing, Article ID 71459, 14 (2006)
9. Lucchese, L., Cortelazzo, G.M.: A noise-robust frequency domain technique for estimating planar roto-translations. IEEE Trans. Signal Process. 48(6), 1769–1786 (2000)
10. Keren, D., Peleg, S., Brada, R.: Image sequence enhancement using sub-pixel displacement. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 742–746 (1988)
11. Irani, M., Peleg, S.: Super resolution from image sequences. In: Proc. IEEE Int. Conf. Pattern Recognition, vol. 2, pp. 115–120 (1990)
12. Irani, M., Peleg, S.: Improving resolution by image registration. Graphical Models and Image Processing 53(3), 231–239 (1991)

13. Chatterjee, P., Mukherjee, S., Chaudhuri, S., Seetharaman, G.: Application of Papoulis-Gerchberg method in image super-resolution and inpainting. Comput. J. 52(1), 80–89 (2009)
14. Zomet, A., Rav-Acha, A., Peleg, S.: Robust super-resolution. In: Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 645–650 (2001)
15. Pham, T.Q., van Vliet, L.J., Schutte, K.: Robust fusion of irregularly sampled data using adaptive normalized convolution. EURASIP Journal on Applied Signal Processing, Article ID 83268, 12 p. (2006)
16. 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)
17. Li, C., Bovik, A.: Three-component weighted Structural Similarity index. In: Proc. SPIE. Image Quality and System Performance VI, vol. 7242, p. 72420Q (2009)
18. Khursheed, K., Imran, M., Ahmad, N., O'Nils, M.: Bi-level video codec for machine vision embedded applications. Elektronika Ir Elektrotechnika 19(8), 93–96 (2013)
19. Fawcett, T.: An introduction to ROC analysis. Pattern Recognition Letters 27(8), 861–874 (2006)

Image Processing with Process Migration

Zdzisław Onderka and Dariusz Półchłopek

Dept. of Geoinformatics and Applied Computer Science,
AGH, Univ. of Science and Technology, Krakow, Poland
zonderka@agh.edu.pl, polchlopek.dariusz@interia.pl

Abstract. In this paper the preemptive migration for the image processing is presented. The migration is defined based on the information policy, transfer policy and location policy. The tests for the image negative calculation were performed. In case of big images minimum 4000×4000 pixels the speedup 2.5 to 3 was achieved.

1 Introduction

The networks of heterogeneous workstations are the standard computing environment form calculations of the Grand Challenge Problems [1]. For example, it could be any CAE (Computer Aided Engineering) application for the large scale problem [9]. Also, a lot of calculations related to the image processing requires large computing power so one can apply for the execution of the calculations, heterogeneous distributed environment.

The real time of the sequential execution of such computation can be very long and is often enlarged by the frequent disk transmitions while the RAM memory is not large enough for storing the data. It could be caused in case of too small size of computer RAM at all or in case of the other applications execution at the same time on the same machine, which need huge memory for their own calculations. Moreover, the users sometimes required the peek of computing power on a single workstation.

The computing power of the computer network is under-utilized most of the time [2]. A consequence of the features mentioned above is utilization of the dynamic load sharing mechanisms [7] i.e. if there are computers with very low load and free resources (like a big amount of RAM memory), they can be used to relieve over-utilized workstations. It could be achieved by the preemptive process migration [6]. In such migration the execution of the task is suspended on the current machine and then resumed on the other one [6]. In [4] was described the Object Oriented aproach to the design of the task migration platform in the heterogeneous computer network. The load sharing problem consisting of an information policy, a location policy and a transfer policy was presented and the Migration Software Development Kit is derfined. This MSDK platform was implemented using Java programming language and to migrate tasks in the netwerk was used CORBA communication standard. One of the major advantages of the CORBA standard is the the ability to implement distributed objects using different programming languages (e.g. C/C++, Java, Pyton) and the possibility of

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_19

using the various system platforms in contrast to the standards dedicated for specific programming languages (like RMI, EJB) or system environments (like .NET Remoting, COM/DCOM, WCF) [5].

In this paper the Object Oriented approach to the preemptive process migration is presented for the image processing. The migrated process is an object, whose execution could be stopped, serialized on the current machine and then sent to and deserialized on the other one and there resumed. Additionaly, the OO techniques enables to define the base classes which can be used to define the inherited classes of objects for the user-specific calculations.

2 Migration Algorithm

The migration algorithm must be based on well-defined load sharing policy. As defined in [6][7] load sharing algorithm is based on three main parts: an information policy, a transfer policy and a location policy. The information policy specifies which information is used in deciding a process migration and where this information is available in the system. The transfer policy defines the need to migrate for each process and the location policy defines target migration host for each task.

2.1 The Information Policy

To improve system performance and also reduce the calculation time we must choose the best alternative from computers in LAN. To accomplish this, we should use the agent system which would have to monit the actual CPU, RAM memory and hard drive use. Based on this data system should have to predict a forecast use of computer resources and choose the best computer in LAN to finish the calculation. IP address of computers in LAN must be loaded to system in a startup of program and should be easy to modify.

2.2 The Transfer Policy

Transfer policy is performed automatically. At the time slow down the calculation, the system automatically starts the migration process. Thus, we can very quickly move the calculation to another computer without user intervention. Transfer between computers was implemented using a TCP-IP protocol which provides end-to-end connectivity and specify how data should be formatted, addressed, transmitted, routed and received at the destination. .NET Framework has two classes for TCP connection - TCPClient and TCPListener. TCPClient initialize connection to another computer, which is waiting using TCPListener class. When client connect to server occurs sending file in 1024 byte packages. If the sending is over, the connection between client and server is closed.

2.3 The Location Policy

When the system decides that the process must be migrated choose the best possible computer and start migration algorithm. The implemented application defines arbitrally the target komputer to which the calculations will be migrated. However, in practice selection can be based on the agent system, which will monit and predict load of computers in a heterogeneous network [3]

3 Project and Implementation

3.1 Functional Requirements

As an example of the image processing was selected task of calculating the negative of any image. After loading the image which will be processed it is displayed in the main window of the serwer application. Algorithm starts on geting value of first pixel in RGB and subtract it from 255. The result of equation must be saved in the same pixel replaceing primary value. In our application we focused on big images minimum 4000×4000 pixels. For smaller images migration makes no sense.

It was assumed that at any time of the calculation, another process is started, which absorbs most of the computer resources (CPU, memory). Immediately after starts this process the time of the calculation increases, which causes run the process of transferring algorithm. The first step is to stop the calculation algorithm for data serialization time. It allows you to save in binary file the state of the object. If you have any problems with the application (for example absorbing process will take to much computer resources causing suspend of computer work), we can start it again and load and then deserialized file which will allow us to get the status of the object before closing and continue the calculation.

The second step is to remotely start the process. The problem was implemented using RMI. To accomplish this you need the access client session data i.e. login, password and IP address. All settings should be placed in the configuration file located in the same directory that the applications executable file. Sent instruction opens the same application with the parameters regarding the amount of bytes to be transmitted, and the IP address of the server. Thanks to this, when client finish the calculation, he would be know where send data back. Then the file is sent to the client. This is done via TCP-IP, which provides us with data integrity and prevents problems with the connection using .NET class (TCPClient, TCPListener). Transfer operations operate on port 8000.

After this operation, server computer comes back to the calculation starting with the point where finished. Server also runs a thread, which monitoring LAN for completed data. Migration file include object of serializable class which contains loaded and partly calculated image and number of amount converted pixcels. This number is very important, becouse it shows the client point to start calculation. Object is serialized to binary file using .NET serialization. When the client computer finish the calculation serialize the effect of his and server work

to a binary file. Then he send back the file to the server, which deserialize it and show the result of calculations.

It may also be the case that the process of absorbing resources lasted so short that the server end faster calculations than the client. In this case, the server does not receive the file from the client, because there is no need. Client computer after receiving the file will automatically deserialize it to form the object and thus saved values accede to the calculations in the place where the server ended the calculation. This can significantly speed up the calculation, without having to calculate everything over.

3.2 Object Diagram

The application consist five classes presented at the Fig. 1. MainForm class is responsible for the appearance of the application. Primary class is the serialization class. Object of this class is transferred to another computer. It contains processed image and state of all variables in class. All information about client and server computer are stored in Access class. The Serialize class contains methods to serialize and deserialize objects. All communication between server and client is implemented in Connection class.

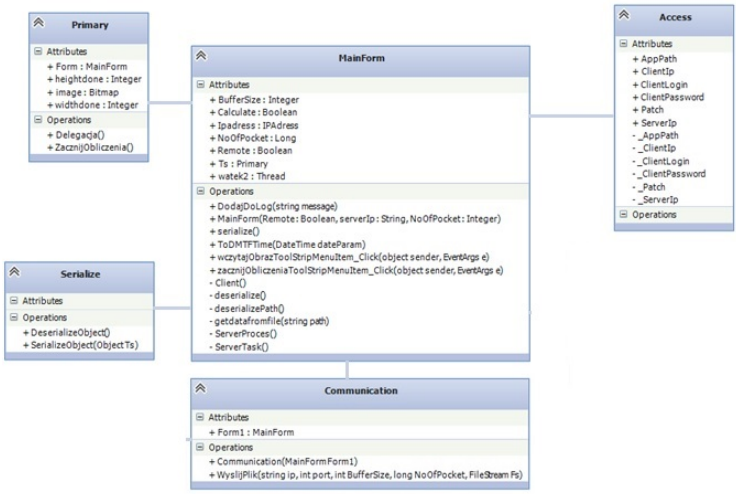


Fig. 1. Objects diagram for the migration of the image processing

3.3 Sequence Diagram

On the Fig. 2 was presented the sequence diagram described the standard behavior of the application using the migration of the calculations on another computer on the local network (described in the section 3.1).

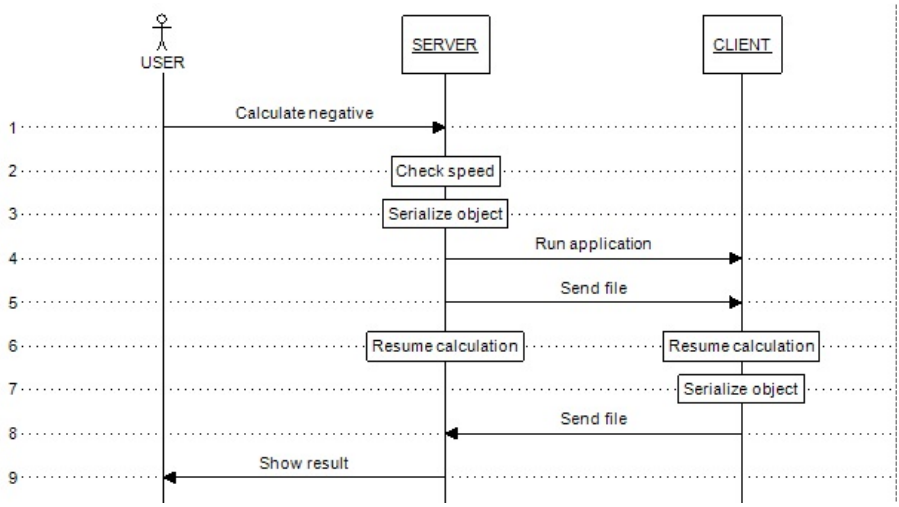


Fig. 2. Sequence diagram for the migration of the image processing

4 Tests

4.1 Environment

The tests were performed on the local network consisting of two computers and one WiFi rooter. First computer worked under the Microsoft Windows 8.1 64 bits operating system (Intel Intel Core i5 3210M @2.5GHz with Turbo Boost mode to 3.1GHz 2 cores, 4 threads, RAM 8GB, Hard Drive SAMSUNG SSD 840EVO 120GB, Graphics Card NVIDIA GeForce 630M, 96 CUDA cores, network card Intel Centrino Wireles-N 2200) and the second one under Microsoft Windows 7 ulimate x64 (Intel Core i5 3230M @2.6GHz with Turo Boost mode to 3.2GHz 2 cores 4 threads, RAM 4GB, Hard Drive TOSHIBA 500GB, Graphics Card INTEL HD Graphics 4000 with 16 execution units up to 332.8 GFLOPS at 1300 MHz, network card Atheros AR946x). The management of the wireless network is realized by the broadband router TP LINK TL-WR740N which serves the wireless 802.11n standard with maximum data transfer rate 150 Mb/s.

4.2 Implementation

Application was implemented using C# programming language on the .NET platform. The following libraries were used:

- System.Net.Sockets - to communicate between two computers in LAN,
- System.Threading - to start new thread, which will check if the client finished calculations,
- System.Runtime.Serialization - to save class object into a file.

To simulate the increased load on the computer system were implemented a program consisting of 4 threads. In every thread there are initialized arrays 10000000 element length of type double and write to them values 0-9 999 999. This program absorbing computer resources was written in C# programming language on the .NET platform.

The essence of the test is the time to calculate the negative of the loaded image. Takes into consideration was the total time from the start of calculation until their completion, compared to the expected length of calculations from which was subtracted the time of serialization and communications between computers.

Absorbing process always starts after runnig the calculations of negative image. For the system recognized the need for migration the time of calculations for 4 consecutive columns of negative image must be greater than the average value of the previous column plus one of third time the calculation of the column.

For the tests was used the following configuration file:

```
serverIP&192.168.0.12
clientIP&192.168.0.13
clientLogin&darek
clientHaslo&darek
FilePath&C:/temp/Migracja/odbior.dat
AppPath&C:/Users/darek/migracja/MigracjaProcesow.exe
```

On the Table 1 were presented the results of image processing for three kinds of images. On every kind of resolution we check three different image, to be sure that our calculations are correct. In order to obtain quantifiable results for each image were performd 10 tests and the Table 1 consists of average values of measured times.

Table 1. The results of execution times with and without migration

Image resolution	Comm. Time	Time to serialization	Percent to serialization	Time of client calculation and send back file	Total time with migration	Total time without migraton
7459×4315	2s	10s	17%	45s	57s	158s
7459×4315	2s	18s	28%	44s	64s	115s
7459×4315	2s	11s	19%	43s	56s	122s
5968×4223	3s	5s	11%	51s	59s	194s
5968×4223	2s	8s	21%	36s	46s	151s
5968×4223	3s	12s	21%	37s	52s	131s
21438×177726	10s	81s	13%	602s	693s	1761s
21438×177726	11s	35	5%	565s	611s	2137s
21438×177726	11s	60s	9%	598s	669s	1596s

First column in table represents the file in pixels. Next column describes required time to send serialized object as file from one computer to another. The third column show the time from the beginning of calculation to serialization. How many percent of entire calculation server done is visible on column four.In the fifth column there's a data how much time elapsed since client start calculation until the data is received by the server. The last but one columns represent the total time with migration. The last one describes thetotal time of entire operation without migration time plus prediction finish time from server.

The communication time takes between 3.125% and 4.34% of total time with migration for the images 7459×7459 and 5968×4315 and about 1.6% of total time with migration. Most of the tests resulted in speedup of the order of 2.5 to 3.

5 Conclusion

The presented above study, tested the image processing example - negative calculations - with the migration of the calculations. The migration is utilized in case of increase of computer load and is realized on local WiFi network consisting with only two computers. Based on performed test it is clear that the time of calculation with migration is significantly lower than without migration. That's why every issue, not just a negative image can be solved quickly with the help of migration of process.

Implemented application runs on the local network only composed of two computers, however, you can easily make an application for extension module to interact with the agent system tested current load of the computers in the network and generating forecast such a load. In this case, such agent system will be an important element of the location policy.

References

1. Agarwal, R.K.: Parallel Computers and Large Problems in Industry. In: Proc. Computational Methods in Applied Sciences. Elsevier Science Publisher (1992)
2. Theimer, M.M., Lantz, K.A.: Finding Idle Machines in a Workstation-Based Distributed System. IEEE Trans. on Software Engineering 15(11) (1989)
3. Lepiarz, M., Onderka, Z.: Agent System for Load Monitoring of The Heterogenous Computer Network. In: Wyrzykowski, R., Dongarra, J., Paprzycki, M., Waśniewski, J. (eds.) PPAM 2001. LNCS, vol. 2328, pp. 364–368. Springer, Heidelberg (2002)
4. Onderka, Z., Uhruski, P.: The SDK for the Process Migration in the Heterogeneous Computer Networks. Schedae Informatica 11, 99–114 (2002)
5. Onderka, Z.: DCOM and CORBA Efficiency in the Wireless Network. In: Kwiecień, A., Gaj, P., Stera, P. (eds.) CN 2012. CCIS, vol. 291, pp. 448–458. Springer, Heidelberg (2012)
6. Alard, E., Bernard, G.: Preemtive Process Migration in Networks of Unix Workstations, pp. 2–4. Inst. National des Telecomm. EVERY, France (1992)
7. Bernard, G., Steve, D., Simatic, M.: A Survay of Load Sharing in Networks of Workstations. Distrib. Syst. Engng 1, 75–86 (1993)

8. Onderka, Z.: Stochastic Control of the Scalable High Performance Distributed Computations. In: Wyrzykowski, R., Dongarra, J., Karczewski, K., Waśniewski, J. (eds.) PPAM 2011, Part II. LNCS, vol. 7204, pp. 181–190. Springer, Heidelberg (2012)
9. Onderka, Z., Schaefer, R.: Markov Chain Based Management of Large Scale Distributed Computations of Earthen Dam Leakages. In: Palma, J.M.L.M., Dongarra, J. (eds.) VECPAR 1996. LNCS, vol. 1215, pp. 49–64. Springer, Heidelberg (1997)

Automatic Extraction of Business Logic from Digital Documents

Paweł Pieniążek

Faculty of ETI, Gdansk University of Technology,
Gabriela Narutowicza 11/12 80-233 Gdańsk, Poland
pawel@pieniazek.eu

Abstract. Executable papers have been attracting attention of scientific publishers, as they may improve validation of their content by reviewers. The paper reports on the ongoing research on making regular PDF file submissions executable without a need to build sophisticated Web systems and allowing authors to use their customary authoring tools. A unified approach has been developed to enable extraction of useful information from the digital PDF document, either scanned printouts or born-digital content and illustrated with experiments involving BPMN diagrams.

1 Introduction

One of the key features of an executable paper is the ability of its reader to freely access various data specified by its content. These may involve *data series* represented in the form of function graphs or table columns, data or activity *flows* represented by block or flow diagrams, music represented as complex text structures, and so on. In order to provide a meaningful information to the reader, data must be interpreted with some predefined interpretation patterns. The Interactive Open Document Architecture IODA [3], has introduced the concept of a multilayered document with separate *data* and *information* layers. The data layer contains the original document file, e.g. PDF, data extracted from it and the executable code for processing that data, whereas the information layer specifies the respective patterns for interpreting them. In the case of scientific papers the most common patterns would be tables, function graphs and flow diagrams.

Practically any typesetting or authoring tool can interchangeably generate PDF (e-paper) documents, as well as print them directly on paper. The PDF format is universal and information is displayed in the same way on any device. A PDF document is visually perfect, however its file content lacks a logical structure, making it unparseable to computers.

In the paper we distinguish two main types of PDF content: purely digital and born digital. The former are just scans of documents printed on paper. Tables, graphs and diagrams are represented in the PDF as graphic raster objects (streams), whereas the latter preserve in their content some amount of vector data information, along with some information on the vendor and tool used

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_20

to generate it, like *GnuPlot*, *tikZ*, *Microsoft Word*, *LibreOffice Writer*, among others.

The main concept of the paper is to enable extraction of useful information from the PDF paper regardless of its provenance, i.e. either scanned or generated. Data can be extracted from PDF binary streams by applying OCRs or other specialized tools, e.g. a handwritten block diagrams recognizer [4], or by converted from vector data in some parseable format, e.g. SVG produced by *Inkscape* [].

Our earlier work focused on extraction of information from tables [2] in born-digital documents and block diagrams from purely digital documents [4]. Throughout the rest of this paper we focus on extracting workflow diagrams from both purely and born digital documents. Extracted diagrams will be expressed in a machine readable form, suitable for workflow editors and workflow simulation engines.

2 Graphical Representation of Business Process

Business Process Model and Notation is used to present business process in a graphical way using a set of primitives. BPMN diagrams can be generated by dedicated tools, such as *TWE* or *Yaoqiang*, which can generate machine readable lexical structures. Another possibility is to draw a diagram with general authoring tool, like Visio or PowerPoint or typeset with Latex using tikZ, amoung others, in the latter case, no logical structure would be created, only the graphics alone, but the document could also be printed.

There is a finite set of primitives in BPMN, which are the following.:

- **Swimlanes** There are two types of swimlanes in BPMN: Pools and Lanes.
- **Tasks** Task is a single atomic action that is executed during the business process. Task may occur in several possible types: basic, send or received message, user, service and script. Tasks are represented by a rectangle with rounded corners. Distinguishing feature, intended for types, is a small pictogram in the upper left corner.
- **Blocks** Block defines subprocess, which is a grouping component of a diagram. Blocks look the same as tasks, but the difference is that they have some other elements inside it.
- **Gateways** he main purpose of gateway is to split or join the workflow. There are two basic types of gateways: Parallel and Exclusive.
- **Events** BPMN allows three kinds of events: *start*, *intermediate* and *end*. Each could be on one of following types: send messages, receive messages, error, link and some other. Difference in representation between start and end is in edge counter thickness. A start event is marked with a thin line, while the end with a thick one. Intermediate events are marked with double lines at the edges.
- **Connections** BPMN provides 3 types of connections: *transitions* (normal lines) – show basic, synchronous flow; *message flows* (dashed lines) – show asynchronous message flow; *associations* (dotted lines) – show the connection between artifacts and other elements.

3 Content Extraction

PDF is a very specific data format including graphical objects scattered all over the document file. In consequence it is hard to read content at the specific page, without recreating it graphical view. Fortunately, there are some useful tools supporting that. For instance *PDFtools* gives the ability to split document to pages or simply crop a page fragment. Another helpful tool is *Inkspcape*, which allows to convert vector parts of the PDF document to SVG data format. On the other hand *Ghostscript* provides the ability to save PDF as a raster image.

The basic concept of extracting information from digital documents is outlines in Fig. 1. Extraction can be divided into several phases. There are two possible input formats: PNG or SVG. The first phase is recognition of the raw graphics, which could be vector or raster, to detect primitives. Recall from the previous section that, it may involve recognition of rounded rectangles, circles and squares. Either type of input file is handled differently. For raster images there is image recognition is involved, whereas for vector images, a parser is used. From that point, both types of file are processed in the same way: first, logic is analysed, next text is recognized and finally XPDL file is generated.

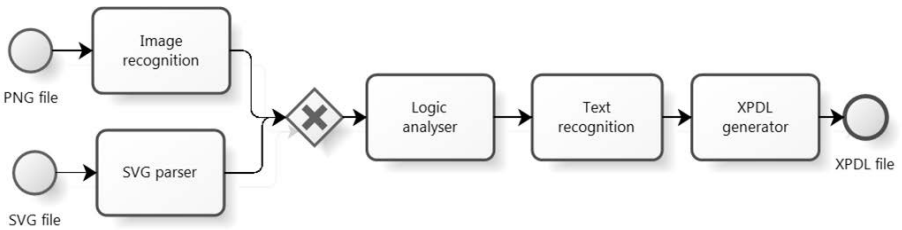


Fig. 1. Content extraction phases

3.1 Image Recognition

We assume that images contain properly drawn diagrams, which are complete, not skewed, with perpendicular lines and without errors. The image recognition phase is divided into stages. The first stage of image recognition is pre-processing. It includes two substages: thresholding the image and skeleton detecting by using Pavlidis algorithm. Mainly this algorithm is for counter tracing, but can be used for thinning the objects which result in ridge detection.

After running this algorithm, all lines are 1-pixel wide. Additionally, all small objects including dots are removed from the picture. In consequence dotted lines disappear, as for the first few steps they are not necessary. Arrows at line ends are emponensily neglected also depredated to lines, and all connections loose information about direction. They will be restored at the later stages. Another problem with this stage is the continuity of the line that has not been preserved in the skeleton crossing points. It is necessary to use algorithm, which restore line continuity, to fix it.

After that, image recognition enters the *detection* phase. Each detected object is removed from the picture to facilitate further analysis. The first stage after preprocessing use histogram analysis. It allows to detect the longest horizontal and vertical lines, which are part of *pools* and *lanes*. Another step is to detect *tasks* and *blocks*. The first step is to detect long line segments - histogram data are used. As it was mentioned before, they are both represented by rounded rectangles, so it contain 4 line segments which has ending which transform into rounded corner. In contrast to it, connection line segments have perpendicular line endings, which can be easily detected. Chosen rounded segments are lengthened in both ways. The places where those new created lines intersect each other are marked as corners. Knowing the line extension direction, type of corner can be estimated. Four different corners, arranged in a rectangular shape, are marked as a rounded rectangle. This can be either block or task; they will be distinguished in *Logic Analyser*.

The next stage is a circle detection. The best way to do that is to use Hough Transform extension, which use the equation of a circle. The first substage looks for circles with radius from 15-30 pixels-that will be events. Second substage focuses on smaller circles, 4-8 pixels. That will be message flow line beginning. This stage is a little bit insidious, due to the fact that the algorithm finds only perfect circles, with radius specified as integer values. After rasterization and skeletization, initial circles can be distorted. This causes necessity of certain assumptions. While at the circle detection stage, it is important to check circle width on initial image to determine event kind.

Another stage is to detect gateways. They consists of four parallel lines, which are situated at 45 degrees angle in relation to whole diagram. Those lines can be detected by using basic Hough Transform algorithm. Further proceedings are the same as in rounded rectangles.

Next stage, which is a line segment detection, consists of three substages. At first the long horizontal and vertical lines are detected. Lines with such parameters dominate in the image and for further detection of additional lines, those detected lines should be deleted. Second stage is based on calculating Hough Transform every time a new segment is detected and deleted. Angular lines are also possible to find in this stage. The stage ends when no more segments with minimum required length can be found. Last stage is use to detect dotted and dashed lines. Each dash and dot are lengthened to merge with each other. After that the fourth stage starts. It is identical like second one, but this time segments for message flows and associations are detected.

As it was said in a previous section, each activity is represented by a shape with pictogram, which denotes type. There are only a few types of pictogram so the analysis is based on simple conditions.

3.2 SVG Parser

Each authoring tool that can generate PDF, produces vector graphics in its specific way [2]. The subsequent conversion to SVG preservers those differences,

which makes it necessary to create heuristics based on specificity of authoring tools.

Upper part of Fig. 2 demonstrate two similar BPMN fragments of diagrams. Both are taken from the same PDF document [1], each one created with a different authoring tool - GNU plot and LibreOffice Draw, respectively. At the first glance, they look identical. The difference can be found after zooming in. These differences are important when creating heuristics for automatic recognition.

For example, a dotted line generated by LibreOffice Draw is shown as a collection of very short line segments, while its GnuPlot counterpart is displayed as a line with dotted style (Fig. 2). Clearly, pattern matching algorithms, used for recognizing SVG, would not find the same pattern in both diagrams. Presented example shows the scale of the problem - just two tools are significantly different in representation of vector data!

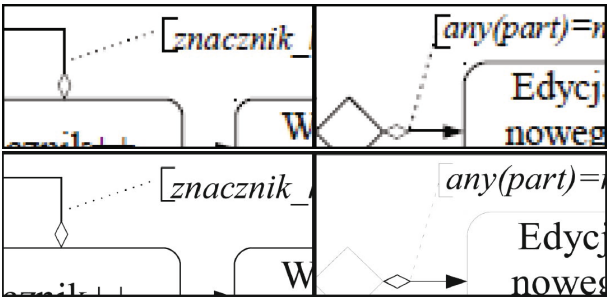


Fig. 2. Comparison between Draw (left) and GnuPlot (right) - image unscaled in upper part; image scaled 1600% in lower part

There were a lot of heuristics developed for LibreOffice Draw, for example: *"L C L C L C L C Z"* means *block* or *task*. The SVG file, converted from PDF does not have any information about primitive shapes. Everything is as simple as possible, but fortunately the letters are not written as simple line segments.

Created SVG file contains only three main types of element: image, text, path. Image and text are quite obvious. Text has attribute named Content, which contains raw text and image is stored as its binary representation. Path can contain nine types of commands. The most commonly used commands are: **M** (moveTo) - change the start drawing point; **L** (lineTo) - draw a line from actual point to the specified point; **C** (curveTo) - draw a BÃľzier curve using corresponding points; **Z** (closePath) - draw a connection between first and last point of the path.

3.3 Logic Analysis

Earlier phases of recognition detect only primitive shapes. To distinguish rounded rectangle between *task* and *block*, additional analysis must be performed. Each

element is checked for overlapping any other one. If element overlaps one another, it means that the larger contains the smaller one. If the element contains at least one element it will be marked as a *block*.

The logic analysis links single line segments together and converts them into polylines. The whole polyline connection must detect source activity, destination activity and distinguished type of connection. To achieve that, each connection must be directionally lengthened. In the nearest area of those lengthened line parts, ending and starting type of arrowhead can be detected. Type of arrowhead allows to define what type of connection line it could be. This phase distinguish all types of connections. The next step is to lengthen linesâĂŹ part again. New part almost certainly will overlap one of the activities and will determine connection source and target. First the transitions and message flows are examined, whereas association are analysed at the end, due to the fact that they can refer to other connections i.e. transitions. The last part of logic analysing is to attach some events to Task. This takes place when the Intermediate event overlap any task activity.

3.4 Text Matcher

Only when the estimated position of the text in the image is known (inside *tasks* and in the *pool* and *lane* headers), the text recognition stage can begin. There are two possibilities depending on the input file format: an OCR tool or text stripping from the SVG tree. PDF tools mentioned before can handle almost all problems with information extraction, however the precise text retrieval is not solved. This is because the authoring tools use various advanced algorithms for justifying text. Different character kerning values (spaces between characters) for each character are applied to achieve visually better results, not only by justifying separate letters to individual objects, but also vertically written text. In the latter case, each letter is separately transformed.

Before merging character into bigger pieces of text, letters must be matched with corresponding elements. Almost any text is placed inside elements to which it is referring, so the position clearly defines its parent element. The only problem is with comments, which has elements limited in only three sides. The width of element is unknown, but it may be estimated.

Changing single characters to text is not straightforward, because each detected character has its precisely defined location. Sorting these locations along the horizontal axis, and then along vertical axis can arrange letters in a correct order. Merging them together creates a single piece of text. The only problem is to detect spaces between words. They are font style dependant and must be calculated using the characterâĂŹs width extracted from the document font dictionary.

3.5 XPDL Generator

The resulting effect of the recognition process is a XPDL document. `XML Process Definition Language` is a language format that describes process flow. It has a

hierarchical structure and stores three types of data. Basic data which stores info about Processes, Activities and Connections. Furthermore, graphic and logic info can also be added in XPDL. Basic data structure of XPDL document is quite simple. Each element genre and kind has its own XPDL tag. Types of elements are stored in attributes. Lines have information about source and destination. Elements from sub processes (blocks) are defined separately to elements from main process.

The XPDL structure is not complex, but that is big problem with generating such files. Each BPMN tool that allows graphical edition, needs a slightly different data structure. There is no universal way to create these files. In the research reported in this paper, three XPDL tools have been tested:

Together workflow editor (TWE – it requires additional graphic information attributes (characteristic only for this tool) to show each type of element. Basic information is sufficient to identify genre, but not for types and kinds. TWE also has problems with some BPMN primitives, e.g. Intermediate Events.

Yaoqiang – requires advanced business logic information. Everything must logically interact. For example: it does not support empty task, with no implementation information or several outgoing connections without restriction tag.

4 Conclusion

For the purpose of this paper, a dedicated software application has been developed. It can accept two types of input format: SVG and PNG, assuming that input images contain only BPMN diagrams. Diagrams are extracted from a PDF file using additional tools [2]. For example, Fig. 3 shows the input image and output XPDL file, graphically interpreted by one of the XPDL editing tools tested during the research.

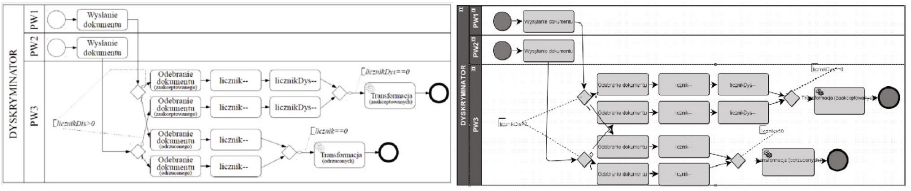


Fig. 3. The input diagram extracted in a PDF (left) and extracted XPDL file interpreted by the in Yaoqiang tool (right)

It is worth emphasizing The main conclusion is that SVG parsing is not as easy as it might seem to be, despite of its generally structured form, SVG trees had to be analyzed, almost like with OCR. Both pure digital and born digital

documents required heuristics to cope with a huge variety of input data and various preferences of vendors.

It is planned, in the nearest future, to create heuristic for each major authoring tools and their major versions will be created. Fortunately, PDF stores information about a vendor and version of authoring tool. For the unknown vendor, raster image processing would be applied. Future work would cover topics about recognition of different types of diagrams specified by UML.

Acknowledgments. This work is supported by the National Science Centre, Poland, under grant DEC1-2011/01/B/ST6/06500.

References

1. Godlewska, M.K.: A model of an open architecture of distributed electronic documents supporting decision processes in collaborative organizations. PhD Thesis, Gdansk University of Technology, Gdansk, Poland (November 2013)
2. Siciarek, J.: Semantics driven table understanding in born-digital documents. In: Choras, R.S. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 149–156. Springer, Heidelberg (2014)
3. Siciarek, J., Wiszniewski, B.: IODA - an interactive open document architecture. In: Proceedings of the International Conference on Computational Science, ICCS 2011. Procedia Computer Science, vol. 4, pp. 668–677 (2011)
4. Szwoch, W., Mucha, M.: Recognition of hand drawn flowcharts. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 4. AISC, vol. 184, pp. 63–70. Springer, Heidelberg (2013)

Comparison of Assessment Regularity Methods Dedicated to Isotropic Cells Structures Analysis

Adam Piórkowski[1], Przemysław Mazurek[2], and Jolanta Gronkowska–Serafin[3]

[1] Department of Geoinfomatics and Applied Computer Science
AGH University of Science and Technology,
A. Mickiewicza 30 Av., 30059 Cracow, Poland
pioro@agh.edu.pl
[2] Department of Signal Processing and Multimedia Engineering,
West–Pomeranian University of Technology, Szczecin,
26. Kwietnia 10 St., 71126 Szczecin, Poland
przemyslaw.mazurek@zut.edu.pl
[3] Department of Ophthalmology
Pomeranian Medical University
Powstańców Wielkopolskich 72 Av., 70111 Szczecin, Poland

Abstract. Assessment of regularity based on quantitative measurements is important for biological structures analysis. The most important is the analysis of the cells' structures. In this paper a few assessment methods are compared: hexagonality, coefficient of variation of sizes, and average coefficient of variation of cells' sides lengths (the new proposed by the authors). This evaluation is based on synthetically generated image datasets for unbiased Monte Carlo analysis. The cell structures are time–evolved using optimization approach starting from irregular to regular.

1 Introduction

Cell structures are recognized especially in the context of complex biological objects, where multiple living cells create geometrical structure in 2D or 3D space. Estimation of the parameters of such a structure is important for the analysis of biological properties. Structures depend on the age, environmental conditions or preparation parameters, so it is very important for medicine and biology. Special interest is related to the living cells, where the observation of the time dependency, relying on the mentioned influence factors could be possible. Analysis of the evolution of such structures is very important for biological and medical researches and applications.

Reliable estimators that are dedicated to the particular cell structures, from the measurements point of view are necessary. Investigating of the estimators of structures, cannot be based on the availability of samples only. A numerical model of the structure that is related to the organization criteria could be proposed as a solution of this problem. The comparison and validation of different estimators could be possible, using the low–cost computations. Selected estimators that are the most promising could be applied to the available image databases.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_21

2 Cell Structures Synthesis and Regularity Evolution

Cell structures could be generated synthetically (Fig. 1) using pseudo random number generator for 2D image plane. The generated points are assumed as a centers of cells. The Dirichlet tessellation algorithm is applied for the synthesis of cells boundaries [1,7]. Such technique creates cells with piecewise linear boundaries. This is image is stretched in specific direction for forcing the isotropic behavior.

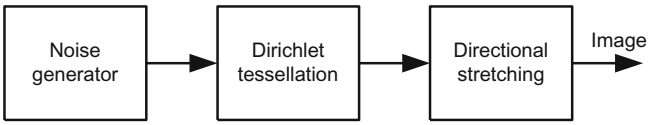


Fig. 1. Image synthesis schematic

Examples of anisotropic and isotropic cell structures are shown in Fig. 2. The strength of stretching could be modified by extending of cells length.

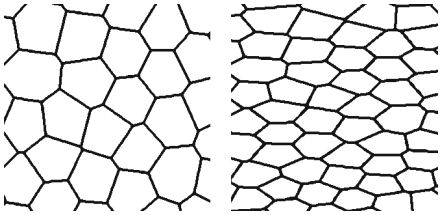


Fig. 2. Examples of anisotropic (left) and isotropic (right) cell structures

The proposed algorithm for the image synthesis is dedicated to single configuration of cells. Testing of the regularity requires comparison of different cell structures that are effects of evolution of structure from irregular to regular cases. The additional process for generation of ordered set of images is necessary.

The proposed process is based on the optimization, based on the distance equalization between two neighboring cells. The regular structure before stretching is the hexagonal or square lattice, with equal distance of cells centers of neighborhood cells. The optimization process is based on the reduction of the largest and extension of the smallest distances. New position of cell center P_i is calculated using the following update formula:

$$P_i \leftarrow P_i + \alpha u_i \left(P_i - P_j\right) \tag{1}$$

where the $(P_i - P_j)$ vector is the movement and $u_i \in \langle 0 - 1 \rangle$ is the random value generated for every iteration. The P_j is the nearest neighbor of P_i, and the α is the step strength coefficient.

Overall optimization process is based on the pseudo random number generator so obtained results are also random variable. It means that two cells structures are less or more regular in mean sense. Particular realization of two iterations (optimization steps) could be better or worse locally. Overall process after hundreds or thousands of iterations is the local or global minimum with regular or almost regular cell structure.

The proposed approach could be used many times for synthesis of cells sets database. All or selected elements (images) from databases could be applied for the comparison of known or proposed regularity metrics.

3 Regularity Methods for Isotropic Endothelium Cells Structures

There are a few attempts to describe a cell's grid in specular microscopy images of corneal endothelial cells [11,2,13]. There are two main factors used: hexagonality and CV. Index of H [4] determines the percentage share of hexagonal cells among all cells in the image (2).

$$H = \frac{N_6}{N_T} 100\% \tag{2}$$

where N_6 is the number of six–side cells, and N_T is the total number of cells in the image.

CV (Coefficient of Variation of sizes) is a factor of relative variance cell sizes [3]. Its value can be calculated from the formula (3).

$$CV = \frac{1}{\mu_c} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (c_i - \mu_c)^2} \tag{3}$$

where c_i is area of i–th cell, μ_c is a mean of size of all cells in the image.

These two factors are insensitive for stretching the shapes. It was proved on a model grids [5]. Therefore, the new factor was proposed [5], $CVSL$. The Coefficient of Variation of the Cell Sides Lengths for each cell is given by the formula (4).

$$CVCSL = \frac{1}{\mu_{SL}} \sqrt{\frac{1}{N_L} \sum_{i=1}^{N_L} (l_i - \mu_{SL})^2} \tag{4}$$

where l_i is the length of i–th side, μ_{SL} is a average length of all cell sides and N_L is the number of sides of a given cell.

To refer to the entire image, the average value ($CVSL$) of factor $CVCSL$ for all cells (N) is calculated (5).

$$CVSL = \overline{CVCSL} = \frac{1}{N} \sum_{j=1}^{N} CVCSL_j \tag{5}$$

The method of determining cell sides' lengths is more detailed described in [5]. To assess a correct values a proper segmentation is needed [6,10,12].

4 Verification of Correlation for Proposed Factor

To validate a quality of proposed factors, the experimental test were carried out. The test was processed using dedicated software tool for the analysis of images of endothelium cells structures. The screen image of developed software is shown in Fig. 3.

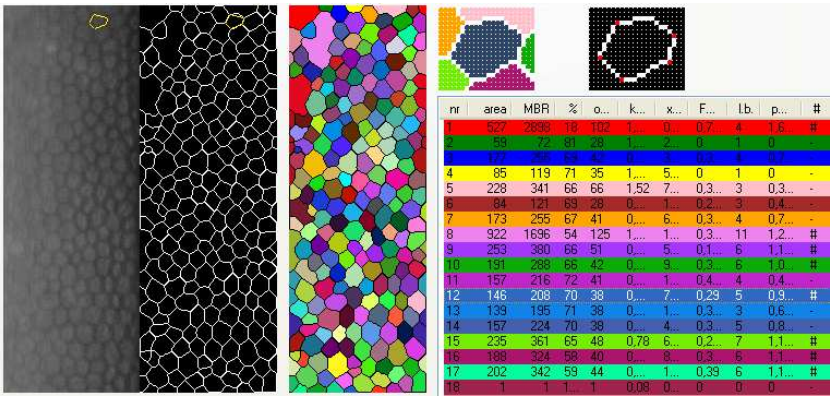


Fig. 3. Screen image of endothelium cells structure analysis software

The tests consisted in assessing the value of pointers for the grid of cells generated artificially in a manner previously described. There were values of proposed factors calculated for a set of 300 successive stages from a disordered state to a state to a state with the largest ordering (Fig. 4). In addition, it was examined the way of indicators behave when the grid cells would extend along one of the axes (AX, AY), or both at once - the scaling process (Fig. 5).

The obtained results (specific coefficients) are analysed further using Matlab software.

The results obtained are shown in the graphs respectively for the CV, H and $CVSL$ (Fig. 6). The coefficients CV and H are almost unchanged for grids stretched and for scaled. In the other way - the $CVSL$ factor was very sensitive for changes in case the grid was stretched in one direction, but kept the value in the case of a uniform scaling of grid. This means that the sensitivity of $CVSL$ factor occurs in the case of stretching, and at the same time the lack of sensitivity in case of scaling.

The slope of curve depends on the generator and all results are shown in Fig. 6. The iteration number cannot be used as important factor for the analysis of results and the validation of coefficients. It is results of the regularity steps necessary to deliver changes of the structure. The validation of the particular coefficient could be based on the observation of the mean value of the coefficient over iteration. The cell structure generator is based on the random number generator applied for the regularity evolution. This generator is main source of

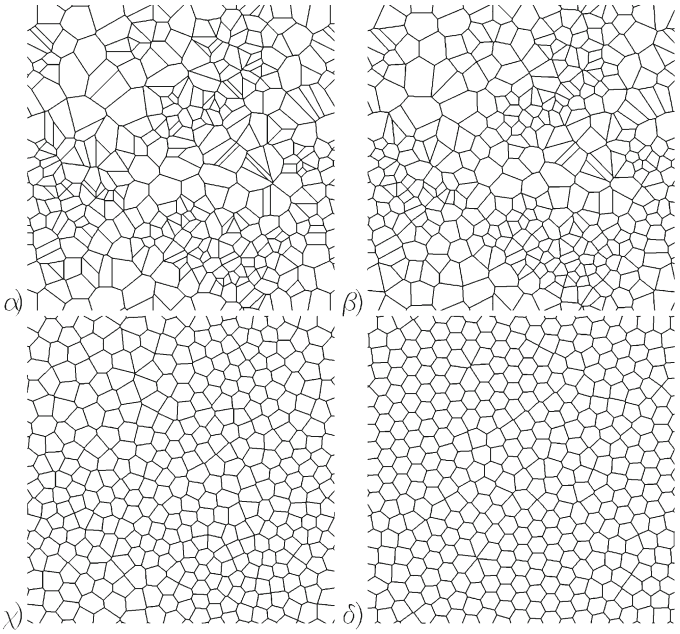


Fig. 4. Artificial grid of cells, stages a) 1, b) 20, c) 100, d) 300

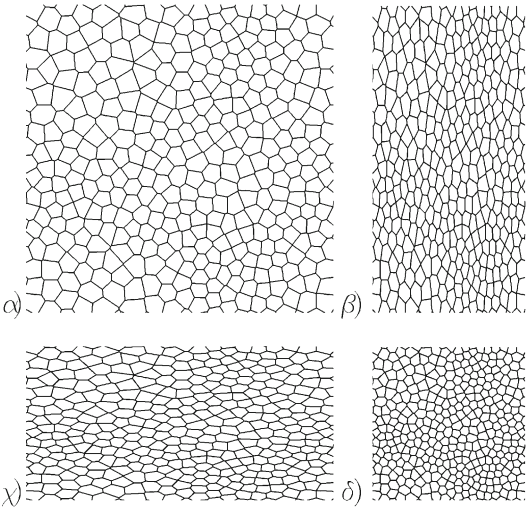


Fig. 5. Cells for the stage 50: a) normal, b) stretched AX 50%, c) stretched AY 50%, d) stretched AX and AY 50% (scaled)

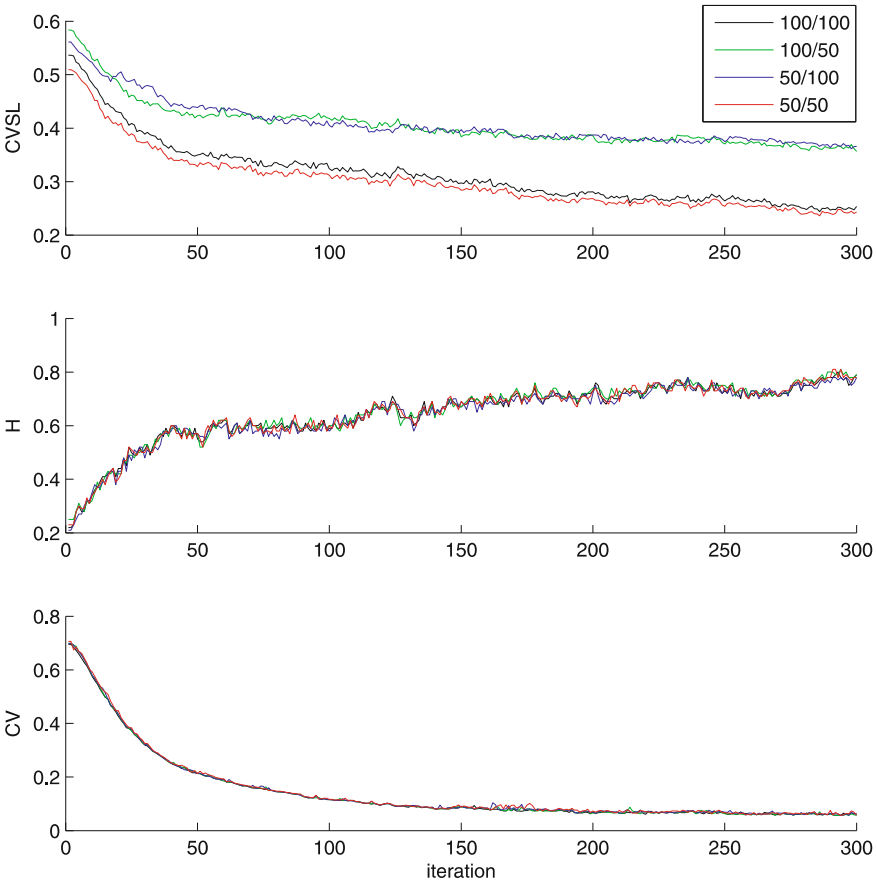


Fig. 6. The values of CV, H and $CVSL$ factor for test series

the random observation of values in Fig. 6 what is expected. This behavior is not a disadvantage of this method, but it is important for the local sensitivity analysis.

The small fluctuations of the cell structure should not be observed as high value peaks. The lack of peaks is the important characteristic of the analysed coefficients using local Monte Carlo search. It means finally, that H, CV and $CVSL$ are proper coefficient not sensitive on the small changes of the structure. It is shown for original and scaled structures together.

The isotropic analysis should be verified using correlation. Strong positive relationship is obtained for H values, and the correlation is shown in Fig. 7. The number of particular cases influence on the 2D histogram value, so it could be influence of the iterations and the generator on the result (the color of the 2D histogram cell). Such influence is not important, because the linear and almost exact relation between H value pairs are obtained.

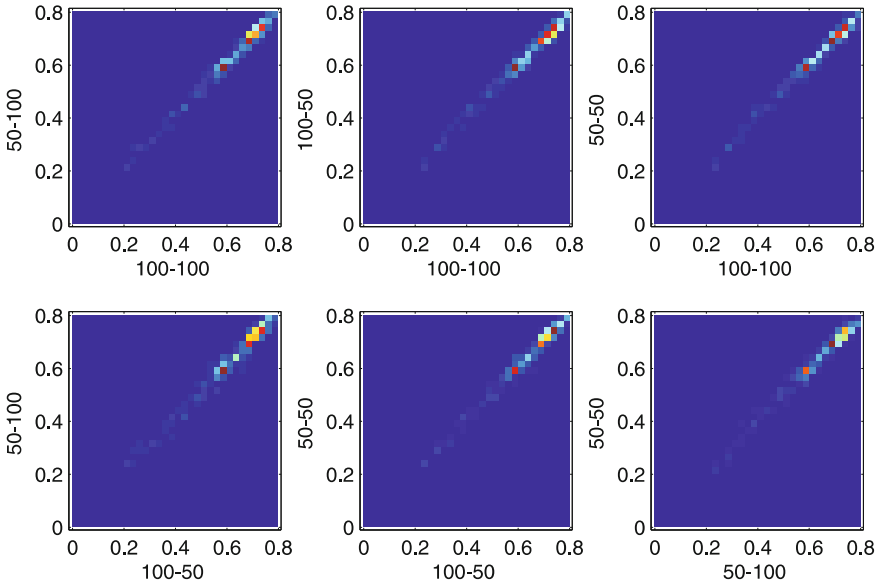


Fig. 7. Correlation of H for test series

Similar strong correlation is obtained for CV values depicted in Fig. 8 and has similar behavior to H. The correlation that is obtained for $CVSL$ values is depicted in Fig. 9 and is different.

The shape of correlation proves relations that could be observed in Fig. 6. Additional line is added for the better visualization of correlation.

5 Discussion and Conclusions

This paper presents a consideration on factors used for endothelium cells' grid quality describing. There are taken into account two standard factors: Hexagonality (H) and Coefficient of Variation of cells' sizes (CV). The new factor - average Coefficient of Variation of the cell Sides Lengths (CVSL) is considered. The tests carried out for a synthetic grid. The results show, that the new factor is sensitive for stretching the grid or cells, unlike the two common factors used for corneal endothelium describing.

Example lack of the sensitivity Hexagonality and CV on stretching [5] is depicted in Fig. 7 and Fig. 8 for large data set. This behavior is important in specific cases of image analyses, when such behavior is desired. The differences that are related to the stretching (anisotropy of the cell structure) should be emphasized using different factor, like proposed CVSL. Analysis of the time dependent changes of the structure related to the different strength of stretching is desired. Measurements of this process using isotropy dependent factor is desired.

CVSL value for particular cell structure is higher for the anisotropic case (50/100) or (100/50). Isotropic cases (100/100) and (50/50) gives two times lower

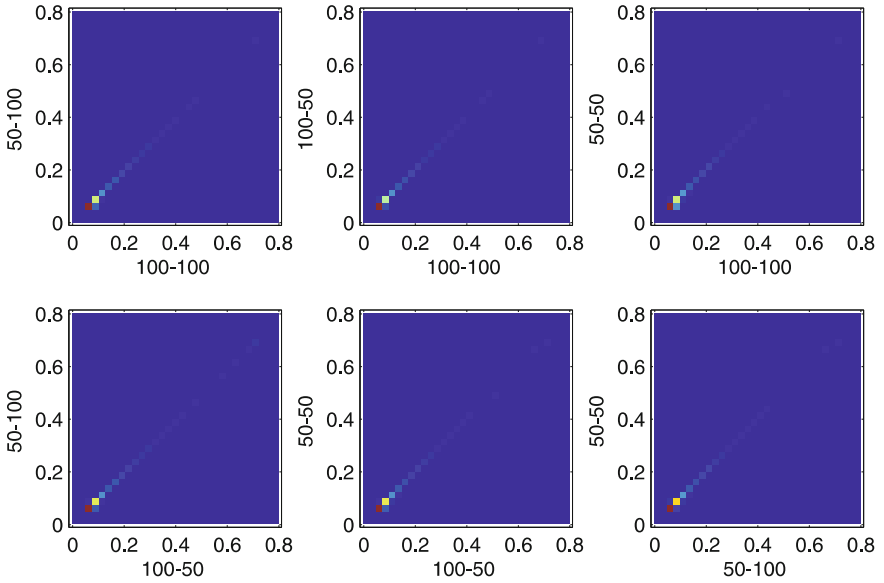


Fig. 8. Correlation of CV for test series

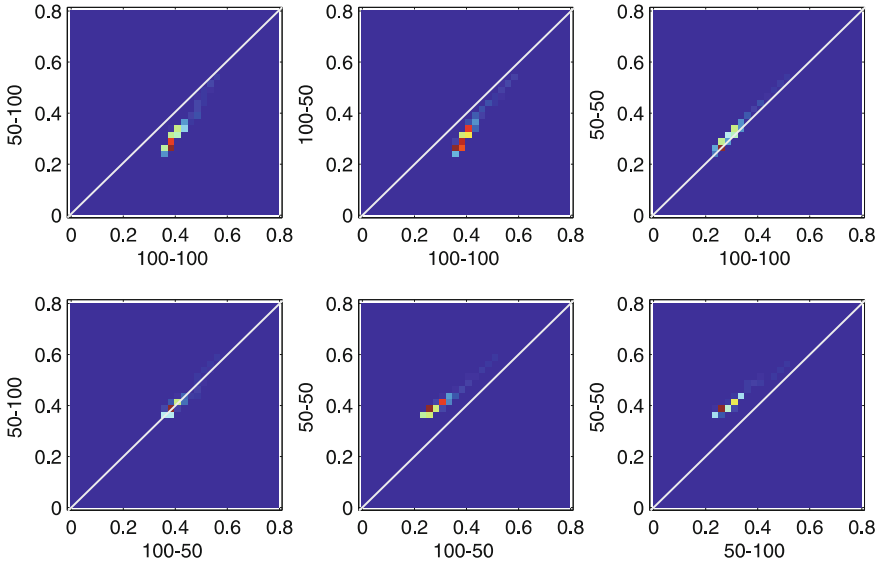


Fig. 9. Correlation of $CVSL$ for test series

values. The Correlation between values obtained for particular cell structures occurs for isotropic stretching. Comparison of isotropic and anisotropic case gives lower correlation what is desired. Pair values are located on the line, but not on the exact correlation line (depicted as a white line in Fig. 9).

Further work considers on the use of the proposed pointer with data from real examinations. The main part of the application for corneal endothelium cells analyzing has been finished and should be released in the nearest future. The multiple synthetically generated tests using Monte Carlo approach will be used for the validation of the relation between anisotropy strengths. Such test are time consuming, because single cell structure should be tested for different stretching. Multiple cell structures should be tested, also.

Very interesting approach for the cell structure analysis is the multi–fractal analysis using e.g. variogram [9] and Triangular Prism Method [8].

Acknowledgments. This work was financed by the AGH - University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, Department of Geoinformatics and Applied Computer Science as a part of statutory project.

This work is supported by software and hardware of the UE EFRR ZPORR project Z/2.32/I/1.3.1/267/05 "Szczecin University of Technology – Research and Education Center of Modern Multimedia Technologies" (Poland).

References

1. Aurenhammer, F., Klein, R.: Voronoi Diagrams. In: Handbook of Computational Geometry, pp. 201–290. Elsevier (2000)
2. Ayala, G., Diaz, M., Martinez-Costa, L.: Granulometric moments and corneal endothelium status. Pattern Recognition 34(6), 1219–1227 (2001)
3. Doughty, M.: The ambiguous coefficient of variation: Polymegethism of the corneal endothelium and central corneal thickness. International Contact Lens Clinic 17(9-10), 240–248 (1990)
4. Doughty, M.: Concerning the symmetry of the 'hexagonal' cells of the corneal endothelium. Experimental Eye Research 55(1), 145–154 (1992)
5. Gronkowska-Serafin, J., Piórkowski, A.: Corneal endothelial grid structure factor based on coefficient of variation of the cell sides lengths. In: Choras, R.S. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 13–20. Springer, Heidelberg (2014)
6. Oblak, E., Doughty, M., Oblak, L.: A semi-automated assessment of cell size and shape in monolayers, with optional adjustment for the cell-cell border width-application to human corneal endothelium. Tissue and Cell 34(4), 283–295 (2002)
7. Okabe, A., Boots, B., Sugihara, K., Chiu, S.: Spatial Tessellations - concepts and Applications of Voronoi Diagrams. John Willey (2000)
8. Oszutowska-Mazurek, D., Mazurek, P., Sycz, K., Waker-Wójciuk, G.z.: Estimation of fractal dimension according to optical density of cell nuclei in Papanicolaou smears. In: Piętka, E., Kawa, J. (eds.) ITIB 2012. LNCS, vol. 7339, pp. 456–463. Springer, Heidelberg (2012)
9. Oszutowska–Mazurek, D., Mazurek, P., Sycz, K., Waker–Wójciuk, G.z.: Variogram Based Estimator of Fractal Dimension for the Analysis of Cell Nuclei from the Papanicolaou Smears. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 4. AISC, vol. 184, pp. 47–54. Springer, Heidelberg (2013)

10. Placzek, B.: Rough sets in identification of cellular automata for medical image processing. Journal of Medical Informatics and Technologies 22, 161–168 (2013) ISSN: 1642-6037
11. Rao, G.N., Lohman, L., Aquavella, J.: Cell size-shape relationships in corneal endothelium. Investigative Ophthalmology and Visual Science 22(2), 271–274 (1982)
12. Sanchez-Marin, F.: Automatic segmentation of contours of corneal cells. Computers in Biology and Medicine 29(4), 243–258 (1999)
13. Zapter, V., Martinez-Costa, L., Ayala, G.: A granulometric analysis of specular microscopy images of human corneal endothelia. Computer Vision and Image Understanding 97(3), 297–314 (2005)

Towards Automated Cell Segmentation in Corneal Endothelium Images

Adam Piórkowski[1] and Jolanta Gronkowska-Serafin[2,3]

[1] Department of Geoinfomatics and Applied Computer Science
AGH University of Science and Technology,
A. Mickiewicza 30 Av., 30–059 Cracow, Poland
pioro@agh.edu.pl
[2] Department of Ophthalmology, Pomeranian Medical University
Powstańców Wielkopolskich 72 Av., 70–111 Szczecin, Poland
[3] Oejenafdelingen, Regionshospitalet Holstebro
Laegaardvej 12a, 7500 Holstebro, Denmark

Abstract. This article addresses the problem of corneal endothelium image segmentation. The aim is an objective determination of boundaries between cells. This problem has not been solved yet as a fully automatic process, the majority of commercial software requires additional correction by an ophthalmologist.

In the paper there are described two approaches allowing to achieve the segmentation of cells that perform more accurate than standard implementations of Watershed algorithms. There is also proposed an algorithm to improve the existing cells division, based on matching proposed segmentation grid lines to the input image content.

1 Introduction

Cornea is a translucent frontal part of the eyeball and a very important part of the eye's refractive system. Cornea is responsible for 2/3 of the whole refractive power of the eye in human. It consists of five layers: outer epithelium, Bowman's membrane, stroma, Descemet's membrane and inner endothelium.

Corneal translucency is achieved by especial, mainly parallel configuration of stromal fibres. Stable corneal hydratation level (ca. 70%) is one of the most important factors responsible for corneal translucency. Water comes freely (free diffusion) into the corneal structure from anterior chamber fluid and has to be actively pumped out by the endothelial cells. Unlike most of the species, human and primates have a very low, clinically insignificant mitotic activity of the endothelial cells. Having ca 13 000 cells/mm^2 during early stages of fetal life, ca 6 500 cells/mm^2 at birth people experience cell loss in a whole lifetime period, to achieve ca 2000 cells/mm^2 at the age of 80. This is a normal course in a healthy, unaffected eye.

Many kind of trauma (iatrogenic or non- iatrogenic) and local and general disease influence the endothelial cells' lifetime. When the endothelial cell decreases to 300 - 500 cells/mm^2, cornea has no possibility to dehydrate and

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_22

irreversible corneal edema occurs, resulting in highly decreased visual acuity requiring corneal transplant.

Corneal endothelium can be assessed in vivo, the most popular tool for assessment a specular microscope. Corneal density and quality can be assessed. An "ideal" corneal endothelium is a monolayer of hexagonal cells of the same size. Practically, cells of different shape and size can be seen. The higher polymorphism (lower percentage of the hexagonal cells) and polymegetism (difference in size between cells) the less is endothelial cell stability.

Normal endothelium should have hexagonality (H) [3] percentage higher then 70% and coefficient of variation ($CV = SD$ of the cell size/ average cell size) [2] of the cell size below 30%. In our opinion these two parameters are not enough to describe cells' variability, so we have proved to find a new quality parameter that refers to the cells regularity.

2 Corneal Endothelium Cells Segmentation

Segmentation of corneal endothelial cells is important for the clinical purposes. On the basis of a cell grid there are carried out basic clinical parameters [16,1,13], such as hexagonality of cells [3], or coefficient of variation of the cell size [2]. Other indicators, such as $CVSL$, proposed in the work [4], require the precisely selected grid cells.

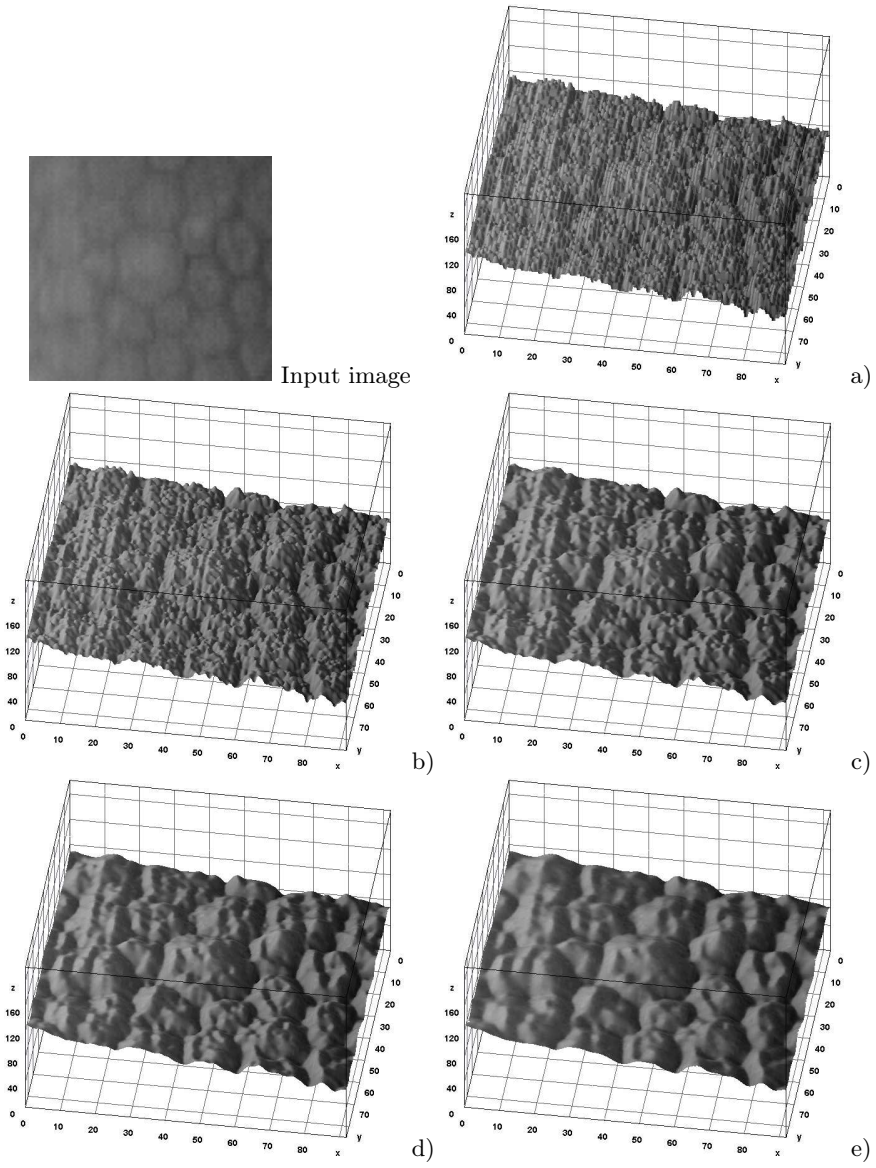
2.1 Problem of Corneal Endothelium Cells Segmentation

Corneal endothelial images presented in this research were taken with non-contact specular microscope Topcon SP-1000. The kind of noise in acquired images does not allow to conduct segmentation using well known algorithms [11,15]. In the Tab. 1 there are shown: the part of an original image and its surface plots - without smoothing and with increasing smoothing (ImageJ - smoothing: 2,4,6,8). It can be noticed that the noise makes a numerous micro hills and valleys, that disturbs the Watershed algorithms. The examples of Watershed processing, implemented in ImageJ program (2 algorithms) are placed in the Tab. 2. There are two cases - without preprocessing a) and with preprocessing d). In the first case there are Watersheds algorithms 1(b) and 2 (c). In the second case a simple convolution with mask 5×5 with '1' was used. The Watershed algorithms' outputs are presented in e) and f). The quality of these segmentations is to poor to assess the clinical parameters. Therefore, the new, objective algorithms are needed.

2.2 Dedicated Methods of Corneal Endothelium Cells Segmentation

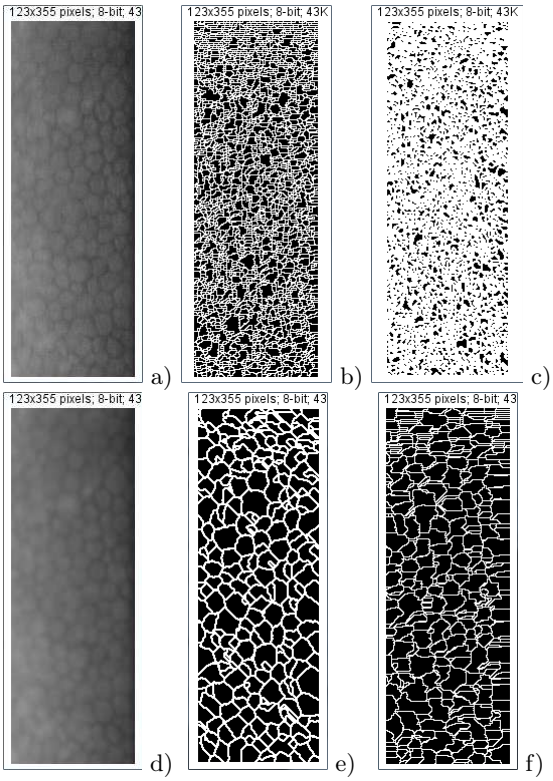
There is a few of approaches for automatic or semi-automatic corneal endothelium cells segmentation. The authors in [6] proposed a non-subsampled Wavelet pyramid decomposition of lowpass region. In the article [14] there is a scale-space filtering used, especially Gaussian, to make use of the separability property of

Table 1. An example and its surface plots for increasing smoothing



Gaussian kernels. A variant of unsharp masking is used too. There are also approaches that are based on classical convolution. In this case there are different masks proposed [7,8,10,5]. These methods do not work properly in some cases. In the Fig. 1 there is presented a sample of corneal endothelium image. The black arrows show the borders of cells, which have a lower contrast than some artifacts

Table 2. An example of segmentation using Watershed algorithms implemented in ImageJ, without (a) and with (d) preprocessing



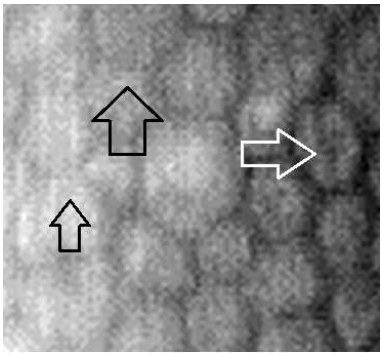


Fig. 1. An example of unsettled contrast of cells boders and inner artifacts

in the cells (the white arrow). Therefore, the method based on the fixed values of masks cannot best mark the cells' borders, so the process of segmentation have to be controlled by human or the output image has to be corrected manually.

3 The Proposals of Corneal Endothelium Cells Segmentation Methods

3.1 Checking the Sign of Changes

The use of convolution filtration has some shortcomings. Masks have fixed values that the clear parts of the image can fit better, but others may mark the artifact as an information (Fig. 1). Hence the proposal to draw from not the difference of the brightness pixel values, but just the sign of these differences. By tracking the valley dividing the cells can determine whether a given point in the context of the neighborhood (eg square mask 3×3) has the smallest number of neighbors with smaller brightness value. This method can set a path leading run-off, in analogy to Watershed algorithms. The calculation of such points in the table n for the image in the table p below describes the algorithm that is shown below.

```
for (x = 1; x < Width -1; x++)
    for (y = 1; y < Height - 1; y++)
        for (i = -1; i <= 1; i++)
            for (j = -1; j <= 1; j++)
                if (p[x + i, y + j] < p[x, y])
                    n[x, y]++;
```

This algorithm explicitly states the boundaries between cells. Additional artifacts can be removed by filtering objects with small size or area. A sample output (normalized) for an corneal endothelial image (Tab. 3, a) is shown in the figure (Tab. 3, b).

3.2 Tracking the Rays from the Hills

The second approach is to trace rays falling from elevated points. For the whole image there are found the points which in the neighborhood (mask 3×3) have the highest (or equal) values. Then, in all directions $(0 - 360^{\circ})$ there are tested rays - each ray is tracked along the profile until it reaches decreasing values (Fig. 2). When there is found a point at which the value increases, the previous point is marked in the final image as a part of boundaries. As an additional parameter can be assumed a minimum radius equal to the average radius of the contour of the cells. This will limit the noise in the image.

At the next stage erosion and dilation are applied to the output image. The result of the segmentation is performed correctly, requiring a manual adjustment of a few places. There is an example in the Tab. 3). For the input image (a) there is the output image (c), next stage of erosion and dilatation (d) and thinning with two different versions of Golay's Alphabet (e, f).

3.3 Improving the Segmentation Output

Segmentation obtained by different methods differ as to the result. This is unacceptable for clinical calculations. Hence the proposal to adjust the segmentation according to the input image before preprocessing.

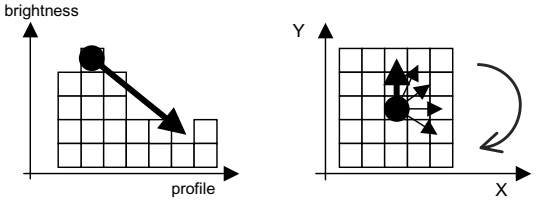
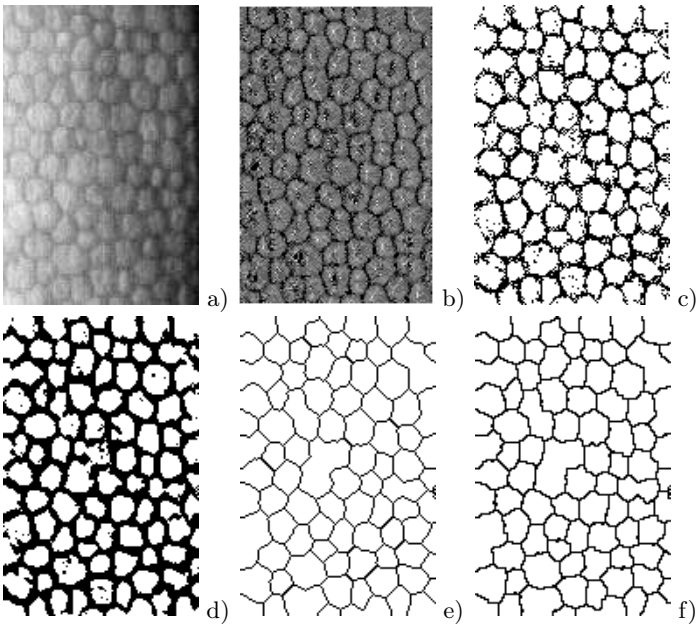


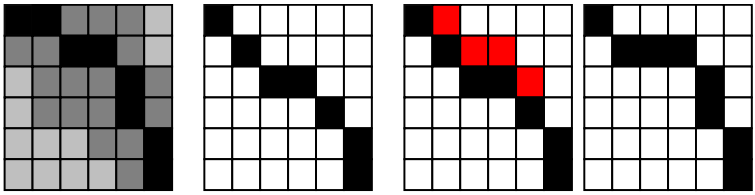
Fig. 2. An idea of tracking the rays

Table 3. A sample corneal endothelial image (after preprocessing) and the outputs of presented algorithms



The concept of the improving algorithm is to examine whether the initial segmentation of cell boundaries run through the points with the lowest local values. To this end, the resulting image is tested whether each boundary point to has no neighbors (in terms of 3×3 masks) with lower brightness in the resulting image than selected (Tab. 4). If so, this new point is added to the segmentation grid. The procedure is performed until the new points are added. After this procedure thinning is started, which should move slightly boundaries. This procedure is repeated until there are changes or a cycle is not detected.

Table 4. An idea of the improving algorithm



4 Conclusions

This article presents the problem of segmentation of corneal endothelium images. It was pointed out that due to the features of noise it is quite difficult to make an objective segmentation. Unshaped boundaries in relation to the internal artifacts in the cells can affect a segmentation as well. There are proposed two algorithms - the first accurately points out the borders in most cases. The second algorithm bases on the content of the image and allows to set limits and boundaries also in questionable places. Both algorithms require a preprocessing, which may influence the final course of boundaries. There is also presented the algorithm to move the boundaries to representative points of the input image to obtain an objective grid of cells' borders. The algorithms can reach the rather large effectiveness, however it is possible to improve the result based on intelligent methods such as neural networks or the rough sets [9,12].

Acknowledgments. This work was financed by the AGH - University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, Department of Geoinformatics and Applied Computer Science as a part of statutory project.

References

1. Ayala, G., Diaz, M., Martinez-Costa, L.: Granulometric moments and corneal endothelium status. Pattern Recognition 34(6), 1219–1227 (2001)
2. Doughty, M.: The ambiguous coefficient of variation: Polymegethism of the corneal endothelium and central corneal thickness. International Contact Lens Clinic 17(9-10) (1990)
3. Doughty, M.: Concerning the symmetry of the 'hexagonal' cells of the corneal endothelium. Experimental Eye Research 55(1), 145–154 (1992)
4. Gronkowska-Serafin, J., Piórkowski, A.: Corneal endothelial grid structure factor based on coefficient of variation of the cell sides lengths. In: Choras, R.S. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 13–20. Springer, Heidelberg (2014)
5. Habrat, K.: Binarization of corneal endothelial digital images. Master's thesis, AGH University of Science and Technology (2012)

6. Khan, M.A.U., Niazi, M.K.K., Khan, M.A., Ibrahim, M.T.: Endothelial cell image enhancement using non-subsampled image pyramid. Information Technology Journal 6(7), 1057–1062 (2007)
7. Mahzoun, M., Okazaki, K., Mitsumoto, H., Kawai, H., Sato, Y., Tamura, S., Kani, K.: Detection and complement of hexagonal borders in corneal endothelial cell image. Medical Imaging Technology 14(1), 56–69 (1996)
8. Oblak, E., Doughty, M., Oblak, L.: A semi-automated assessment of cell size and shape in monolayers, with optional adjustment for the cell-cell border width-application to human corneal endothelium. Tissue and Cell 34(4), 283–295 (2002)
9. Ogiela, M., Tadeusiewicz, R.: Artificial intelligence methods in shape feature analysis of selected organs in medical images. Image Processing and Communications 6, 3–11 (2000)
10. Piorkowski, A., Gronkowska-Serafin, J.: Analysis of corneal endothelial image using classic image processing methods. In: KOWBAN - XVIII The Computer-Aided Scientific Research, The Works of Wroclaw Scientific Society, B, Wroclawskie Towarzystwo Naukowe, vol. 217, pp. 283–290 (2011)
11. Piorkowski, A., Gronkowska-Serafin, J.: Selected issues of corneal endothelial image segmentation. Journal of Medical Informatics and Technologies 17, 239–245 (2011)
12. Placzek, B.: Rough sets in identification of cellular automata for medical image processing. Journal of Medical Informatics and Technologies 22, 161–168 (2013)
13. Rao, G.N., Lohman, N., Aquavella, L., Cell, J.: size-shape relationships in corneal endothelium. Investigative Ophthalmology and Visual Science 22(2), 271–274 (1982)
14. Sanchez-Marin, F.: Automatic segmentation of contours of corneal cells. Computers in Biology and Medicine 29(4), 243–258 (1999)
15. Szostek, K., Gronkowska-Serafin, J., Piorkowski, A.: Problems of corneal endothelial image binarization. Schedae Informaticae 20, 211–218 (2011)
16. Zapter, V., Martinez-Costa, L., Ayala, G.: A granulometric analysis of specular microscopy images of human corneal endothelia. Computer Vision and Image Understanding 97(3), 297–314 (2005)

A Texture-Based Energy for Active Contour Image Segmentation

Daniel Reska, Cezary Boldak, and Marek Kretowski

Faculty of Computer Science
Bialystok University of Technology, Bialystok, Poland
{d.reska,c.boldak,m.kretowski}@pb.edu.pl

Abstract. This paper presents a two-dimensional deformable model-based image segmentation method that integrates texture feature analysis into the model evolution process. Traditionally, the deformable models use edge and intensity-based information as the influencing image forces. Incorporation of the image texture information can increase the methods robustness and application possibilities. The algorithm generates a set of texture feature maps and selects the features that are best suited for the currently segmented region. Then, it incorporates them into the image energies that control the deformation process. Currently, the method uses the Grey Level Co-occurrence Matrix (GLCM) texture features, calculated using hardware acceleration. The preliminary experimental results, compared with outcomes obtained using standard energies, show a clearly visible improvement of the segmentation on images with various texture patterns.

1 Introduction

Efficient and robust image segmentation is one of the main challenges in computer vision. Deformable models [1] are a widely used class of segmentation methods. Generally, a deformable model is an active shape (e.g., a 2D contour or a 3D surface) that tries to adapt to a specific image region. This adaptation process is influenced by external and internal forces, that deform the shape towards the boundaries of the segmented region. The external forces attract the model to desired image features, while the internal forces control its smoothness and continuity. This formulation allows the models to overcome many problems, like image noise and boundary irregularities. External forces can come from a variety of image features. The most common approaches use the edge or intensity statistics of the segmented region [2,3,4]. This makes the deformable models well suited for extraction of areas with distinctive (but not necessary continuous) borders and fairly uniform texture. However, larger patterns with high contrast pose a greater challenge to traditional methods.

Deformable models were introduced with a seminal "snake" algorithm [2], which was a parametric active contour with an edge-based image force and a set of internal energies that controlled bending and stretching of the curve. Since then, the classical model has been heavily modified and extended, e.g., with

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_23

addiction of expansion forces [5], edge-based vector field energies [3], adaptive topology mechanisms [6] and region-based image energies [4,7]. Furthermore, methods combining deformable models with texture extraction and classification [8] were also introduced. Texture-based methods generally fall into two categories: utilising information obtained from supervised texture analysis, or using the calculated features without earlier classification. The supervised methods can require an initial analysis stage, which provides information about the number and characteristics of distinctive patterns in the image [9], or they can use a selected set of features to create a deformation map that influence the evolution of the model [10]. Unsupervised methods can extract the required information using the initial region intensity statistics to create a likelihood deformation map [11], or, in case of patters with a larger scale, use a small bank of Gabor filters to extract the features [12,13]. Texture features can also be used to improve the segmentation process alongside other image characteristics [14].

In this paper, we present a texture-based image energy for a two-dimensional parametric active contour. The energy incorporates texture features into our template active contour framework [15]. The proposed energy can be used alongside other image forces, expanding the application possibilities of the method. Currently, the energy utilises features based on Gray-Level Co-occurrence Matrices (GLCM) [16]. The features are generated using hardware-accelerated implementation, which allows an efficient creation of a large feature set. The preliminary experiments were performed on synthetic images with various texture patterns and show a visible improvement over typical image energies.

2 Texture-Based Active Contour

The proposed algorithm incorporates a texture-based image energy into a discrete parametric snake [17] with an adaptive topology reformulation abilities [6]. The energy makes the snake to expand into a region with an uniform texture that is similar to the initial region of the contour.

Firstly, the snake is manually initialised inside the segmented region. Next, the algorithm generates a set of texture feature maps and selects the features that are best suited for the currently segmented area. Finally, the contour evolution process expands the snake under the influence of the texture energy, which is based on the selected features. This algorithm can be presented in a pseudocode:

Input: Initialisation of the snake $\mathbf{s}$
1: Generate the initial texture feature maps set $\mathbf{T}_{init}$
2: Initialise empty set $\mathbf{T}_{best}$ for selected texture feature maps
3: **for all** texture feature map $\mathbf{t} \in \mathbf{T}_{init}$ **do**
4: **if** $\mathbf{t}$ meets the selection condition **then**
5: add $\mathbf{t}$ to $\mathbf{T}_{best}$
6: **end if**
7: **end for**
8: Create external energy $\mathbf{E}_{tex}$ using $\mathbf{T}_{best}$
9: **repeat**

10: **for all** snake point $\mathbf{p} \in$ contour $\mathbf{s}$ **do**
11: Minimise the local energy of $\mathbf{p}$ using $\mathbf{E}_{tex}$ and other energies
12: **end for**
13: **until** convergence of snake $\mathbf{s}$
14: **return** the segmented region.

The detailed descriptions of the algorithm steps are presented below.

Initialisation. Initialisation of the contour is the crucial step in the current version of our algorithm. The initial snake should be manually placed inside the segmented area and in a region of a possibly uniform texture. Furthermore, the contour area should be large enough to cover the texture pattern and capture the entirety of its characteristics, preferably by including several texture pattern tiles.

Texture feature maps generation and selection. The next step of our algorithm is generation of the texture feature maps. The features are calculated for each pixel of the source image, that results in a set of feature maps.

The currently used texture features, generated from the Grey-Level Co-occurrence Matrix, are: Entropy, Correlation, Homogeneity, Contrast and Energy. The maps are generated for different sets of GLCM parameters: window size (from 3×3 to 11×11 by default), displacement (from 1 to 3 pixels) and orientation (0°, 45°, 90°, 135° and for all four angles). The algorithm, however, is not limited to the GLCM approach – any method that can generate a feature map of the segmented image can be used.

The following algorithm step selects the texture feature maps that will be used in the segmentation process. The algorithm prefers features that have a low dispersion inside the initial contour region. In order to mark the region, a binary mask is created from the initial curve. For the pixels covered by the mask in each of the feature images t_i, the mean of the feature values $\bar{x}_i$, standard deviation σ_i and Relative Standard Deviation $\%RSD_i = \frac{\sigma_i}{\bar{x}_i} \times 100$ are calculated. The texture feature maps used for the segmentation process must have the $\%RSD$ lower than a user-specified threshold (equal to 65% by default). As there is no supervised texture classification step and the number of different textures in the image is not known, this condition selects the texture features that, hopefully, will distinguish the segmented area from regions with different textures.

Additionally, the selection step can reduce the number of maps by analysing their similarity for different orientations in groups with the same feature, window size and displacement. By default, the algorithm generates one map for each of the orientations and one extra map with an averaged response for all angles. Then, it selects only the directional maps which have their feature value mean (inside the initial region) sufficiently different (at least 50% by default) from the feature mean in the averaged map. However, in case of a clearly isotropic texture pattern, generation of the maps with different orientations can be manually turned off and simplified by creating only the averaged response map.

Contour evolution process. The snake evolution process aims to minimise the energies of the contour points (snaxels) by moving them to the positions of the

lowest local energy. A typical intensity energy E^s_{int} for a snaxel s and a potential destination point p can be defined as:

$$E^s_{int}(p) = \begin{cases} \frac{1}{d(s,p)+1} & \text{if } |I(p) - \bar{x}| \leq \theta \times \sigma \\ 1 & \text{otherwise,} \end{cases} \quad (1)$$

where $I(p)$ is the intensity value in point p, $\bar{x}$ and σ are the intensity mean and standard deviation in the snake initial region, θ is the user-specified sensitivity threshold and $d(s,p)$ is the distance between the location of snaxel s and point p. As the energy is inversely proportional to the distance between the current snaxel and its potential position, it prefers more distant points, enabling the expansion of the contour.

Our texture energy E^s_{tex} works in a way similar to the energy described above, but it takes all the generated texture feature map into consideration: the current snaxel can be moved into the new position only if the similarity condition is fulfilled for all the selected texture feature maps, as defined in:

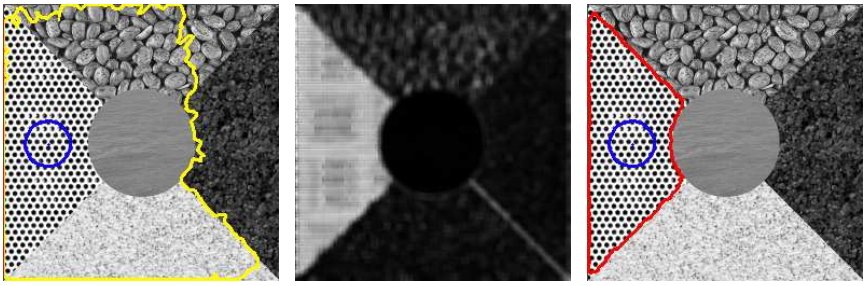
$$E^s_{tex}(p) = \begin{cases} \frac{1}{d(s,p)+1} & \text{if } \forall t \in T_{best} : |val_t(p) - \bar{x}_t| \leq \theta \times \sigma_t \\ 1 & \text{otherwise,} \end{cases} \quad (2)$$

where t is a texture feature map in the selected set T_{best}, $\bar{x}_t$ and σ_t are the feature mean and standard deviation in the snake initial region, $val_t(p)$ is the value of the texture feature in the point p, and θ is a user-defined constant. This energy works under two assumptions: (a) features with a low dispersion in the initial snake region have a potential to discriminate it from other patterns and (b) a significant value change in one of the maps blocks the current snaxel from further movement.

3 Experimental Results

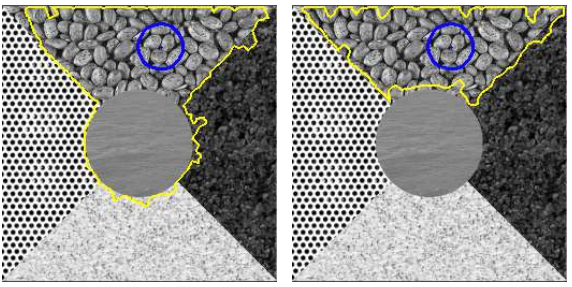
This section shows the segmentation results of the proposed method. The experiments were performed on a machine with AMD FX 8150 Eight-Core processor, 16 GB RAM, Nvidia GeForce GTX 660 graphics card (with 960 CUDA cores), and running on Ubuntu 12.04. The total segmentation time was less than 5 seconds for each of the presented examples.

The algorithm was implemented using the MESA system [15] – a platform for designing and evaluation of the deformable model-based segmentation methods. MESA provides a template system for construction of active contours from exchangeable elements (i.e., models, energies and extensions), allowing an easy comparison of the proposed approach with other energies. The GLCM texture generation algorithm was implemented in OpenCL [18] and integrated with the existing code using a Java binding library (JOCL from www.jocl.org). OpenCL allowed utilisation of the graphical processing units, which effectively led to a significant speedup of the algorithm (from a few minutes on CPU-based implementation to a few seconds using the hardware-accelerated version).



(a) (b) (c)

Fig. 1. Segmentation of a high contrast pattern: (a) result of the segmentation with the default intensity energy (initial circular contour visible in blue), (b) calculated GLCM Contrast map and (c) result of the proposed energy



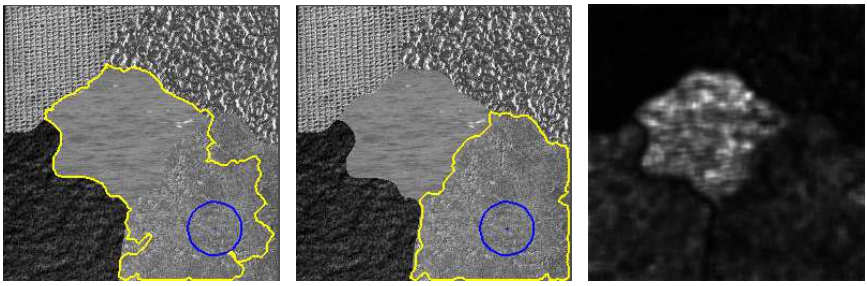
(a) (b)

Fig. 2. Segmentation of a coarse-grained texture: (a) initialisation (in blue) and result (in yellow) using the default intensity energy and (b) result using the proposed energy

The method was tested on 256×256 synthetic images created using the Brodatz texture database [19]. The initial contours were manually placed inside the desired region and scaled to the preferred size. During the experiments, only the sensitivity parameter θ was modified (between 3 and 4), while the other parameters were left constant on default values.

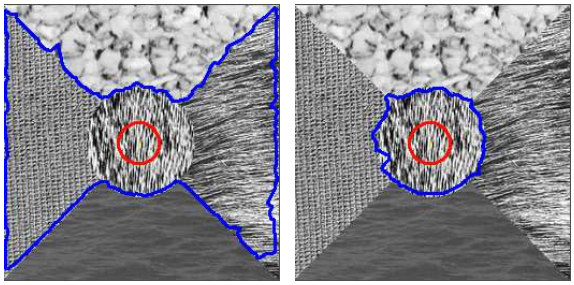
The first example (Fig. 1) presents a segmentation of a region with a high contrast texture. The size and intensity dispersion of the pattern make it impossible to segment with an edge-driven force or with the described default intensity-based energy (Fig. 1a). However, using a GLCM Contrast map (Fig. 1b), the texture-based energy managed to correctly drive the snake to the region boundaries (Fig. 1c).

The second example (Fig. 2) shows the result for a more coarse-grained pattern. Again, the intensity-based energy failed to distinguish between two regions with a similar average intensity. The third example (Fig. 3) shows a difficult situation, where two regions with different textures not only have a similar



(a) (b) (c)

Fig. 3. Example of a fine-grained texture segmentation: (a) result with the intensity energy, (b) result with the texture-based energy and (c) a sample texture map (GLCM Energy) that discriminates the two regions



(a) (b)

Fig. 4. Importance of the texture orientation selection: (a) result without the angle selection and (b) segmentation with orientation selection enabled

average intensity, but also do not have an evident border between them. While the intensity-based energy could not be adjusted to segment the region, the texture-based approach easily separates the two problematic regions. The last example (Fig. 4) highlights the importance of the texture map orientation selection. In a case of a highly directional texture, the energy without angle selection fails to segment the region (Fig. 4a). With this selection enabled, the region is correctly separated from the other textures (Fig. 4b).

4 Conclusions and Future Work

In this paper a texture-based energy for the two-dimensional parametric active contour is presented. This energy improves the segmentation performance on images with textures of various size, contrast and complexity. Moreover, the algorithm does not require any previous information about the texture classes in

the segmented image. Despite the relatively large number of the utilised features, an efficient GPU-accelerated texture generation method enables a high performance of the segmentation process. The currently proposed approach gives promising results on artificially composed images. We are currently investigating the performance and possible application of the method on natural and medical images.

The present form of the algorithm is in an early stage of development and can benefit from many possible improvements. An initial analysis of the start region can be used to find an optimal parameter set for the generated features. Furthermore, the texture feature maps in the initial step can be calculated only for the start region of the snake, while the generation of the complete maps can be performed after the selection, which will improve the performance. Moreover, the currently utilised feature set (based on GLCM) can be easily extended by incorporation of other texture feature extraction methods, like Gabor filters [12].

Acknowledgement. This work was supported by Bialystok University of Technology under Grant S/WI/2/2013.

References

1. Moore, P., Molloy, D.: A survey of computer-based deformable models. In: International Machine Vision and Image Processing Conference, IMVIP 2007, pp. 55–66 (2007)
2. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1(4), 321–331 (1988)
3. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing 7(3), 359–369 (1998)
4. Ronfard, R.: Region-based strategies for active contour models. International Journal of Computer Vision 13(2), 229–251 (1994)
5. Cohen, L.D.: On active contour models and balloons. CVGIP: Image Understanding 53, 211–218 (1991)
6. Mcinerney, T., Terzopoulos, D.: T-snakes: Topology adaptive snakes. Medical Image Analysis, 840–845 (1999)
7. Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing 10(2), 266–277 (2001)
8. Reed, T., DuBuf, J.: A review of recent texture segmentation and feature extraction techniques. CVGIP: Image understanding 57(3), 359–372 (1993)
9. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for supervised texture segmentation. International Journal of Computer Vision 46(3), 223–247 (2002)
10. Pujol, O., Radeva, P.: Texture segmentation by statistical deformable models. International Journal of Image and Graphics 4(03), 433–452 (2004)
11. Shen, T., Zhang, S.G., Huang, J., Huang, X., Metaxas, D.: Integrating shape and texture in 3D deformable models: from Metamorphs to Active Volume Models. In: Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies, pp. 1–31. Springer (2011)
12. Jain, A., Farrokhnia, F.: Unsupervised texture segmentation using Gabor filters. Pattern Recognition 24(12), 1167–1186 (1991)

13. Huang, X., Qian, Z., Huang, R., Metaxas, D.: Deformable-model based textured object segmentation. In: Rangarajan, A., Vemuri, B.C., Yuille, A.L. (eds.) EMMCVPR 2005. LNCS, vol. 3757, pp. 119–135. Springer, Heidelberg (2005)
14. Brox, T., Rousson, M., Deriche, R., Weickert, J.: Colour, texture, and motion in level set based segmentation and tracking. Image and Vision Computing 28(3), 376–390 (2010)
15. Reska, D., Jurczuk, K.F., Boldak, C., Kretowski, M.: MESA: Complete approach for design and evaluation of segmentation methods using real and simulated tomographic images. In: Biocybernetics and Biomedical Engineering (2014)
16. Haralick, R., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics (6), 610–621 (1973)
17. Reska, D., Kretowski, M.: HIST - an application for segmentation of hepatic images. Zeszyty Naukowe Politechniki Bialostockiej. Informatyka 7, 71–93 (2011)
18. Stone, J., Gohara, D., Shi, G.: OpenCL: A parallel programming standard for heterogeneous computing systems. Computing in Science & Engineering 12(3), 66 (2010)
19. Brodatz, P.: Textures: a photographic album for artists and designers. Dover Publications (1966)

Object Localization and Detection Using Variance Filter

Grzegorz Sarwas[1] and Sławomir Skoneczny[2]

[1] LINGARO Sp. z o o
grzegorz.sarwas@lingaro.com
[2] Institute of Control and Industrial Electronics
S.Skoneczny@isep.pw.edu.pl

Abstract. In this paper different variance filters for rejecting image regions that do not contain interesting object are tested. In our case the processed scenes have equally depth of focus, which makes difficult to distinguish objects from the background. In order to locate the object, the algorithm based on the sliding windows approach has been used. In case of using this type of algorithm a cascade of filters designed to reject windows that do not contain searched objects are applied. In this paper the authors put emphasis on elimination of redundant windows, from equally depth colour scenes, using various variance filters. Also a formula, based on the integral images, which can improve the efficiency of using directional variance filters, is proposed. All types of variance filters are tested and compared.

1 Introduction

Most recognition systems are designed in order to find only one object in image or to classify extracted object or scene. They can decide about the existence of the object in the image but they are not able to inform us about its localization.

Nowadays, two main factors play the important role for the automatic scene understanding: objects' localization and relations between them, as well as the background extraction. One of the most successful localization technique is a sliding window method [3]. By using this approach we split the analysed image into all possible sub-windows, which strongly increases the computational burden. There also exist more efficient approaches for subwindow searching problem, e.g. [8,9]. The most advantageous feature of sliding window solution is performed by replacing the localization problem by image retrieval task. It can be solved by using global properties of objects, e.g. colour distribution [11,7,1] and histograms [2,13,4,5], or global statistics of local features, e.g. bag-of-words approach [10]. The main goal of this kind of filter in the detection task is rejecting the most recent regions where is a certain lack of searched pattern.

The problem that we are trying to solve is a multiple finding and classification of all objects from the date base, which exists in the processed colour image. Usually, it is not sufficient to solve a problem of rejecting the subwindows that do not contain target objects, while using only one classifier. Following the detector

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_24

proposed by Kalal in TLD [6] applied to solve this kind of problem, the cascade of filters is necessary. In TLD the first classifier is a variance filter, which can by adopted in efficient way for sliding window approach by using the integral image. While we apply the TLD detector to find only one object the condition for rejecting the subwindow is that its variance is smaller twice than the variance of the target pattern. However, this condition is not sufficient for the scenes that contain many similar objects not distinguishable from the background, with the difference very often based exclusively on the colour (in contrary of the object in shades of grey). The main problem is choosing the variance descriptor that can perform the most precise distinction between colour objects and choosing the rejection threshold for subwindows.

In this paper colour object localization methods using different variants of variance filters are described. In contrast to TLD algorithm we primarily focus on colour distribution. This algorithms are tested on the colour images where is hard to notice the differences between objects and background.

2 Variance Filters

In this section different types of variance filter used in our research are presented. In the performed experiments the authors focus on the RGB colour images, so it is possible to process the intensity of each channel separately. The descriptors used in this process can be divided in two groups:

1. Non-directional descriptors
   (a) Variance of grey scale intensity image;
   (b) Variances of R, G, B channels separately;
2. Directional descriptors
   (a) Variances of means of rows and columns for intensity image;
   (b) Variance of rows and columns variances of intensity image;
   (c) Variances of means of rows and columns all of channels separately;
   (d) Variance of rows and columns variances for all channels separately;

To simplify the non-directional variance an image is considered as a one-dimensional vector of pixels and its elements are addressed using the notation I_i for the i^{th} pixel [6] and variance is defined as: $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(I_i - \mu)^2$, where n is the number of pixels and μ is the mean of the pixels. This formula can also written as: $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} I_i^2 - \mu^2$. Computing the variances for all subwindows is very memory and time consuming task. In next subsections efficient methods solving this problem are described.

2.1 Non Directional Variance

In [12] an efficient method for calculation the variance of image intensity for all subwindows has been presented. This method based on the Integral Image, for which the computation of the pixels sum no longer depends on the number of pixels. However, in this paper Viola made a mistake during derivation of

formula for calculation variance by using Integral Image approach. The correct derivation of this relation can found in [6]. Following the Kalal, a variance of any subwindows $I(B)$ of image I can be calculated using the relation:

$$\sigma^2 = \frac{1}{n} I''(B) - \left[\frac{1}{n} I'(B)\right]^2, \tag{1}$$

where I' and I'' are the integral images of the same size as image I and n is a number of elements.

Definition 1. *I' is an image at location (x, y) containing the sum of all pixel values between the points $(1, 1)$ and (x, y) defined as:*

$$I'(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y').$$

Definition 2. *I'' is an image at location (x, y) containing the sum of the squared pixel values between the points $(1, 1)$ and (x, y) defined as:*

$$I''(x, y) = \sum_{x' \leq x, y' \leq y} I^2(x', y').$$

Following the definition 1 the formula of computing the sum of pixels (or pixels squared) for the subwindow B bounded by coordinates (x, y, w, h) is given by:

$$I'(B) = I'(x-1, y-1) - I'(x+w, y-1) - I'(x-1, y+h) + I'(x+w, y+h). \tag{2}$$

Similarly, we can calculate formula for $I''(B)$:

$$I''(B) = I''(x-1, y-1) - I''(x+w, y-1) - I''(x-1, y+h) + I''(x+w, y+h). \tag{3}$$

2.2 Directional Variance

In [7] Kekre and Patil proposed to use standard deviation of row and column mean vectors comparing with standard deviation of variance. This approach was used for CBIR system for colour images. In this paper we use the variance of row and column mean vector.

Definition 3. *The variance of row mean vector σ_r^2 is defined as:*

$$\sigma_r^2 = \frac{1}{r} \sum_{i=1}^{r} (\mu_i - \mu)^2,$$

where $\mu_i = \frac{1}{c} \sum_{j=1}^{c} I(i, j)$ and μ is a mean of the pixel values:

In case of computing this variance we can use the same integral images like for the standard variance but it is faster to provide the directional integral images I'_r, I''_r, I'_c, I''_c. These integral images are described in Def. 5 and 4.

Definition 4. *I'_c and I''_c are column integral images defined as:*

$$I'_c(x,y) = \sum_{x' \le x} I(x',y), \quad I''_c(x,y) = \sum_{x' \le x} I^2(x',y).$$

Definition 5. *I'_r and I''_r are row integral images defined as:*

$$I'_r(x,y) = \sum_{y' \le y} I(x,y'), \quad I''_r(x,y) = \sum_{y' \le y} I^2(x,y').$$

Then, the formula for computing directional variances can be found.

Theorem 1. *The variance of row mean vector σ_r^2 can by obtained by using the following expression:*

$$\sigma_r^2 = \frac{1}{c^2 r} \sum_{i=1}^{c} I_c^{'2}(B_i) - \left[\frac{1}{rc} I'(B)\right]^2, \tag{4}$$

where r and c denote a number of rows and columns respectively.

Proof. The variance of row mean vector σ_r^2 is:

$$\sigma_r^2 = \frac{1}{r} \sum_{i=1}^{r} (\mu_i - \mu)^2. \tag{5}$$

By transforming variance of row mean we obtain the alternative formula:

$$\begin{aligned}\sigma_r^2 &= \frac{1}{r} \sum_{i=1}^{r} (\mu_i - \mu)^2 = \frac{1}{r} \sum_{i=1}^{r} \mu_i^2 - \frac{2\mu}{r} \sum_{i=1}^{r} \mu_i + \frac{1}{r} \sum_{i=1}^{r} \mu^2 \\ &= \frac{1}{r} \sum_{i=1}^{r} \mu_i^2 - 2\mu^2 + \mu^2 = \frac{1}{r} \sum_{i=1}^{r} \mu_i^2 - \mu^2 = \frac{1}{c^2 r} \sum_{i=1}^{r} \left[\sum_{j=1}^{c} I(i,j)\right]^2 - \mu^2,\end{aligned}$$

which proofs the Theorem 1.

Similarly, we can derive an expression for variance of column mean vector which is defined as:

Definition 6. *The variance of column mean vector σ_c^2 is defined as:*

$$\sigma_c^2 = \frac{1}{c} \sum_{j=1}^{c} (\mu_j - \mu)^2,$$

where $\mu_j = \frac{1}{r} \sum_{i=1}^{r} I(i,j)$ and μ is a mean of the pixel values.

The variance of row mean vector σ_c^2 can by calculated as:

$$\sigma_c^2 = \frac{1}{r^2 c} \sum_{j=1}^{r} I_r^{'2}(B_j) - \left[\frac{1}{rc} I'(B)\right]^2, \tag{6}$$

where r and c are a number of rows and a number of columns respectively.

Next, we define the vectors of row and column variances for subwindow B.

Definition 7. *The vector of row variances* $\mathbf{V}_{\sigma_\mathbf{r}^2}$ *is defined similarly as:*

$$\mathbf{V}_{\sigma_\mathbf{r}^2} = \begin{bmatrix} \sigma_r^2(y) \\ \sigma_r^2(y+1) \\ \vdots \\ \sigma_r^2(y+w) \end{bmatrix} = \frac{1}{h}\mathbf{I}''_\mathbf{c}(\mathbf{B}) - \left[\frac{1}{h}\mathbf{I}'_\mathbf{c}(\mathbf{B})\right]^2,$$

where

$$\mathbf{I}''_\mathbf{c}(\mathbf{B}) = \begin{bmatrix} I''_c(x+h,y) - I''_c(x-1,y) \\ I''_c(x+h,y+1) - I''_c(x-1,y+1) \\ \vdots \\ I''_c(x+h,y+w) - I''_c(x-1,y+w) \end{bmatrix}$$

and

$$\mathbf{I}'_\mathbf{c}(\mathbf{B}) = \begin{bmatrix} I'_c(x+h,y) - I'_c(x-1,y) \\ I'_c(x+h,y+1) - I'_c(x-1,y+1) \\ \vdots \\ I'_c(x+h,y+w) - I'_c(x-1,y+w) \end{bmatrix}.$$

Definition 8. *The vector of column variances* $\mathbf{V}_{\sigma_\mathbf{c}^2}$ *is defined similar:*

$$\mathbf{V}_{\sigma_\mathbf{c}^2} = \begin{bmatrix} \sigma_c^2(x) \\ \sigma_c^2(x+1) \\ \vdots \\ \sigma_c^2(x+h) \end{bmatrix}^T = \frac{1}{w}\mathbf{I}''_\mathbf{r}(\mathbf{B}) - \left[\frac{1}{w}\mathbf{I}'_\mathbf{r}(\mathbf{B})\right]^2.$$

3 Experiments

For experiment the RGB image presented in Fig. 1 was used. This image shows 10 different type of commodities stagnant on a shelves (10 target products). All products in this picture have the same depth of focus. For localization purpose three target products presented in the Fig. 2 have been chosen. The object images have been extracted from the processed image. The first experiment has been performed to check a standard deviation between all instances of target patterns. All objects objects have been manually cut from the processed image. This test shows the divergence between the variances of target products. All RGB channels have been tested separately. The results have been compared with values achieved for grey scale image. Table 1 presents the results of experiments. From the results it follows that objects of this same type have the most similar variance of row and column variance vectors. The second experiment has been aimed to check how many windows we can reject assuming that we want to find all searched products. A condition for subwindow rejection was the variance 0.15 times smaller or higher of the min and max variations for any of the channels of each target objects. For faster and easier analysis we have fixed the subwindow ratio (width/high) to be the same as a ratio of target products. Subwindow shifts

Fig. 1. Processing scene

(a) First object

(b) Second object

(c) Third object

Fig. 2. Localized objects

Table 1. Standard deviations of variances for first and second objects

	First object					Second object				
Channel	σ^2	σ_r^2	σ_c^2	$\sigma^2_{\mathbf{V}_{\sigma_r^2}}$	$\sigma^2_{\mathbf{V}_{\sigma_c^2}}$	σ^2	σ_r^2	σ_c^2	$\sigma^2_{\mathbf{V}_{\sigma_r^2}}$	$\sigma^2_{\mathbf{V}_{\sigma_c^2}}$
Grey	0.0026	0.0007	0.0017	0	0.0001	0.0043	0.0029	0.0002	0.0001	0.0001
R	0.0027	0.0004	0.0019	0	0.0001	0.0061	0.0042	0.0002	0.0002	0.0002
G	0.0027	0.0011	0.0018	0	0	0.0042	0.0029	0.0003	0.0001	0.0001
B	0.0033	0.0005	0.0018	0	0	0.0021	0.0016	0.0003	0.0001	0

Table 2. Number of received windows

Descriptor	Object 1	Object 2	Object 3
Grey non directional	103783	106842	97510
Grey variance of mean	130953	134454	131997
Grey variance of vec	129297	106842	133317
RGB non directional	124361	135554	130768
RGB variance of mean	136218	136498	136315
RGB variance of vec	136271	136447	136368

have been set to 10% of the subwindow sizes. Additionally, the constrains for subwindow size have been introduced. In effect in each test 138355 subwindows have been used. As the result of this test we achieved a number of rejecting windows. The larger value indicates greater efficiency of the filter. These results are presented in Table 2.

(a) RGB variance of variances

(b) RGB variance of means

Fig. 3. Results of second object localization

4 Conclusion

In this paper the different variance filters for receiving subwindows that do not contain targeted objects are tested and discussed. Directional variance of rows and columns have the lowest standard deviation for examined objects. However, the best performance were achieved for RGB variance mean filters, what can be (Fig. 3). Processing in RGB space gave better results than using a pure grey scale model.

While applying directional filters should be remembered that the possibility of their use strongly depends on the position of the target objects in the processed scene. Therefore, we should perform further experiments in order to detect all possible rotation from the principle pattern orientation giving the possibility of detection. Furthermore, we have to test how the colour normalization influences to the colour variances.

Acknowledgement. This work was co-financed by the European Union within the European Regional Development Fund.

References

1. Bhuravarjula, H., Kumar, V.: A novel content based image retrieval using variance color moment. IJCER 1 (2012)
2. Buccafurri, F., Lax, G.: Approximating sliding windows by cyclic tree-like histograms for efficient range queries. Data Knowl. Eng. 69(9), 979–997 (2010)
3. Forsyth, D.A., Ponce, J.: Computer Vision: A Modern Approach. Prentice Hall Professional Technical Reference (2002)
4. Hadjidemetriou, E., Grossberg, M., Nayar, S.: Multiresolution histograms and their use for recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(7), 831–847 (2004)
5. 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)
6. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(7), 1409–1422 (2012)

7. Kekre, H., Patil, K.: Standard deviation of mean and variance of rows and columns of images for cbir. IJCISSE (WASET) 3 (2009)
8. Lampert, C.H., Blaschko, H., Hofmann, M., Beyond, T.: sliding windows: Object localization by efficient subwindow search. In: IEEE Conference on CVPR 2008, pp. 1–8 (2008)
9. Lampert, C.H., Blaschko, H., Hofmann, M., Efficient, T.: subwindow search: A branch and bound framework for object localization. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(12), 2129–2142 (2009)
10. Liu, J.: Image retrieval based on bag-of-words model. CoRR, abs/1304.5168 (2013)
11. Tran, L.V.: Efficient image retrieval with statistical color descriptors. Linkping University, Department of Science and Technology (2003)
12. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Computer Society Conference on CVPR 2001, vol. 1, pp. I–511–I–518 (2001)
13. Wei, Y., Tao, L.: Efficient histogram-based sliding window. In: IEEE Conference on CVPR 2010, pp. 3003–3010 (2010)

The Impact of the Image Feature Detector and Descriptor Choice on Visual SLAM Accuracy

Adam Schmidt and Marek Kraft

Poznan University of Technology, Institute of Control and Information Engineering
Piotrowo 3A, 60-965 Poznan, Poland
adam.schmidt@put.poznan.pl

Abstract. The paper presents an evaluation of a range of contemporary image feature detectors and descriptors for the mobile robot visual simultaneous localization and mapping. The analysis of the impact of the detector and descriptor choice on the accuracy of robot trajectory reconstruction was performed using precise ground truth data. Moreover, as processing time is an important, the average computation times for each of the detector-descriptor pairs are also included.

1 Introduction

Automated detection and matching of point features across a sequence of images is a part of many computer vision algorithms, with robot navigation being no exception. Over the past years, the feature-based approach was commonly used in both the robot visual simultaneous localization and mapping (SLAM) [8] and the visual odometry (VO) [9]. As the quality of the input data (the feature matches) is crucial for the quality of the output of the navigation system, the development of image feature detectors and descriptors is a field of active research.

The evaluation of the performance of various detector-descriptor pairs in the context of the application at hand is highly important, as the characteristics of the features detected with a specific detector can significantly influence the matching process. As far as SLAM goes, to the extent of authors knowledge no such extensive study was performed. Desirable characteristics of keypoint detectors and descriptors were described in [14] and [4], but the authors did not provide any experimental results. The evaluation results presented in [11] and [17] are based on the number of feature matches consistent with the epipolar geometry established between consecutive frames. The choice of such indirect evaluation method was forced by the difficulties with gathering the ground truth feature correspondences and the lack of availability of the ground truth trajectory. An accurate ground truth trajectory was available in the research presented in [3] and [7], as the test sequences were recorded with the camera placed on the industrial robot arm. However, only a limited set of detectors and descriptors was tested. Moreover, using the industrial robot arm constrains the movement range of the camera to the volume of the work envelope.

This papers present an evaluation of the image feature detector-descriptor pairs, including ones recently proposed, in the extended Kalman filter (EKF)

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_25

based visual SLAM framework. The tests were performed on real-world data with accurate ground truth. The measures selected for the evaluation correspond directly to the quality of the trajectory reconstruction. This allows to test not only the robustness of the detection and matching process, but also its impact on SLAM accuracy. Moreover, processing times for each detectordescriptor pair are also provided.

The Section 2 provides a short description of the evaluated detector and descriptor algorithms. The experimental setup, along with the evaluation procedure are presented in Section 3. Results and discussion are given in Section 4, while Section 5 contains the conclusions and presents the directions for future work.

2 Feature Detectors and Descriptors

The section provides a short characteristic of the evaluated algorithms.

Harris and Shi-Tomasi (GFTT) algorithms [10] [18] detect image features based on the analysis of the local image autocorrelation. For each point on the image the structural tensor is computed. The structural tensor is the base for the computation of the corner score functions for both Harris and GFTT. The local maxima of the corner score function with values greater than some arbitrary threshold are marked as image features.

SIFT (Scale Invariant Feature Transform) is the pioneer robust multiscale feature detector and descriptor. The features are found by applying the difference of Gaussians filter at multiple scales and performing a scale space NMS with additional filtering to reject line-like features. The location of the keypoint in the image is interpolated. The final step is the orientation assignment based in the histogram of gradients in the neighborhood of the feature. Computation of the descriptor is performed by dividing the neighborhood into 16 4s×4s subregions (s denotes the scale). In the next step the 8-bin orientation histograms are computed for each subregion. The partial histograms are finally concatenated to form the final, 128-element feature descriptor.

FAST (Features from Accelerated Segment Test) is an algorithm for feature detection [15]. If a continuous segment of n pixels which are either all darker or all brighter than the central pixel by more than the threshold value t is found the pixel is considered a feature candidate. The order in which pixels are tested is was determined using machine learning to achieve high processing speed. The candidates are further refined by using an additional corner score function and applying the non-maximum suppression (NMS).

SURF (Speeded Up Robust Features) algorithm is a multiscale image feature detector and descriptor [5]. The detection step in SURF takes advantage of the use of Haar wavelet approximation of the blob detector based on the Hessian determinant. The approximations of Haar wavelets can be efficiently computed using integral images, regardless of the scale. Accurate localization of multiscale SURF features requires interpolation. The descriptor is also based on the Haar wavelets and encodes the distribution of pixel intensity values in the neighborhood of the detected feature. Computation of the descriptor begins

with the assignment of the dominant orientation to make the descriptor rotation invariant. A square window with a side length of 20s (s being the feature scale) is placed on the feature point and oriented as it was computed in the previous step. The window is divided into 4×4 regular square subregions that are then divided further to 5×5 uniformly distributed sample points. For each sample point, Haar wavelet responses for two principal directions are computed. Each subregion contributes to the descriptor with four components total of 64 descriptor elements.

STAR keypoint detector is a derivative of the CenSurE (Center Surround Extrema) feature detector [1]. It was developed as a multiscale detector with full spatial resolution, as the interpolation performed typically by other multiscale detectors affects the accuracy of feature localization. The detector uses a bi-level approximation of the Laplacian of Gaussians (LoG) filter. The shape of the detector mask was selected to minimize the directional bias to preserve rotational invariance while enabling the use of integral images for efficient computation. Scale-space is constructed without interpolation, by applying masks of different size.

BRIEF (Binary Robust Independent Elementary Features) descriptor proposed in [6] uses binary strings for feature description and subsequent matching. The use of Hamming distance as similarity measure enables very fast computation. As BRIEF is sensitive to noise, an averaging filter is applied to the image before descriptor computation. The value of each bit contributing to the descriptor corresponds to the result of a comparison between the intensity values of two points inside an image segment centered on the currently described feature.

ORB (Oriented FAST and Rotated BRIEF) algorithm for feature detection and description that combines and extends on the concepts of FAST and BRIEF [16]. The feature detection is performed on multiple scales using FAST and augmented the with orientation data to achieve robustness to in-plane rotation. Similarly to BRIEF, a binary descriptor is used, but the coordinates of the point pairs for binary tests around the described feature are rotated by the feature orientation angle. The random sampling used to select the point pairs in BRIEF has been replaced with a sampling scheme that uses machine learning for decorrelating BRIEF features under rotational invariance. This makes the nearest neighbor search during matching less error-prone.

BRISK (Binary Robust Invariant Scalable Keypoints) descriptor is another feature detection and description method based on the FAST and the BRIEF. However, the random sampling pattern was replaced by a regular, circular pattern consisting of 60 points. Each point is used for more than one comparison and the complete descriptor has the length of 512 bits. Neither the scale nor the rotation angle are discretized, which, according to the authors, results in significantly improved robustness.

FREAK (Fast Retina Keypoint) descriptor is based on BRIEF [2]. Unlike BRIEF, the descriptor is orientation aware and uses a different, biologically inspired sampling pattern to select point pairs. The orientation is computed based on the estimated local gradients over selected point pairs in the neighborhood of

Fig. 1. Example frames from the test sequence

the feature. While the resulting descriptor is still a binary string, the sampling pattern allows for the use of a coarse-to-fine approach to feature description. Point pairs carrying the information on most distinctive characteristics of the feature neighborhood are compared in the first place. This allows for faster rejection of false matches and shortening of the computation time.

Additionally, a simple matching method based on normalized cross-correllation (NCC) was included in the evaluation.

3 Experiment Description

The evaluation was performed on a 9.28 meter reference trajectory gathered using the WiFiBot robot platform.

The dataset consists of 800 VGA resolution frames captured using Basler ace GigE Vision cameras. Example frames from the sequence are presented in Fig. 1. The sequence exhibits typical characteristics met during exploration of an unknown environment by a wheeled robot e.g.alternating turns with varying turning radius, forward and backward movement, loop closing, varying lighting conditions, slight motion blur.

The dataset was used to perform trajectory reconstruction using the EKF SLAM framework based on the MonoSLAM approach presented in [8].

The selected images are a part of the PUT RGB-D dataset[1].

The images in the dataset are complemented with high quality ground truth (GT) data. The GT data was gathered using an overhead multi-camera system during the registration of the datatset trajectories.

The absolute trajectory error (ATE) proposed in [19] was used as the accuracy metric for all the experiments. The output of the SLAM system is a series of the estimated positions of the robot $r(i)$, where i denotes the frame number. The measurements are tested against a set of reference positions taken from the GT data $r_{\mathrm{GT}}(i)$. As the monocular systems do not provide the scale information, it was necessary to compute the rotation R, translation t and scale s aligning both

[1] The dataset was downloaded from `http://www.vision.put.poznan.pl/?pageid=28`

the measured and the GT trajectory. With both trajectories aligned, ATE for a point i in the trajectory is calculated as:

$$\mathrm{ATE}(i) = \|r_{\mathrm{GT}}(i) - (sRr(i) + t)\| \tag{1}$$

The final quality of the reconstructed trajectory is the root mean square error (RMSE) of a series of ATE errors for all the trajectory images. The RMSE is calculated as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{ATE}(i))^2} \tag{2}$$

Detector and descriptor implementations were taken from the OpenCV image processing library. The parameters directly influencing the feature matching (e.g. matching threshold, maximum visibility angle etc.) were independently adjusted for each detector-descriptor combination. The combination of the SIFT detector and ORB descriptor was excluded from the evaluation due to the incompatibility of the implementations. As the processing time is crucial for the efficient operation of the visual SLAM, average processing time for all detector-descriptor pairs was also measured. The analysis was performed on a PC with an Intel i7 3.8GHz processor.

4 Results and Discussion

Table 1 contains the RMSE obtained for detector-descriptor combinations while Table 2 contains the maximum ATE for each reconstructed trajectory. The average processing time for each one of the combinations is presented in Table 3.

The best results w.r.t. the RMSE were obtained for the feature points detected with the Harris or the FAST detectors. The system using the SIFT descriptor achieved the best results, contended only by the FAST detector.

Table 1. The RMSE for various feature point detector and descriptor combinations (in meters)

Det. \ Descr.	NCC	SURF	ORB	BRIEF	BRISK	FREAK	SIFT
BRISK	0.147	0.407	0.291	0.203	0.262	0.286	0.358
FAST	0.084	0.274	0.081	0.063	0.089	0.119	0.046
GFTT	0.097	0.716	0.096	0.068	0.082	0.095	0.069
Harris	0.115	0.420	0.070	0.088	0.081	0.138	0.061
ORB	0.186	0.191	0.122	0.108	0.156	0.204	0.316
SIFT	0.642	0.186	-	0.139	0.138	0.077	0.126
STAR	0.330	0.239	0.281	0.156	0.249	0.155	0.109
SURF	0.259	0.184	0.113	0.172	0.133	0.143	0.307

Table 2. The maximal ATE for various feature point detector and descriptor combinations (in meters)

Det. \ Descr.	NCC	SURF	ORB	BRIEF	BRISK	FREAK	SIFT
BRISK	0.311	0.672	0.493	0.526	0.487	0.545	0.579
FAST	0.343	0.518	0.159	0.123	0.190	0.378	0.111
GFTT	0.416	1.350	0.235	0.172	0.280	0.199	0.145
Harris	0.225	0.915	0.127	0.184	0.271	0.254	0.149
ORB	0.402	0.462	0.240	0.219	0.265	0.414	0.740
SIFT	1.278	0.344	-	0.369	0.267	0.175	0.278
STAR	0.691	0.493	0.583	0.284	0.446	0.329	0.264
SURF	0.726	0.463	0.273	0.331	0.210	0.296	0.505

The BRIEF, ORB and BRISK descriptors performed slightly worse. The system using the correlation based matching performed on par with those using the binary descriptors.

The low performance of the systems using the multi-scale detectors probably comes from the fact, that the interpolation of the point features positions introduces localization errors. Moreover, as we are dealing with an indoor scenario, the camera in-plane rotation is negligible and various means of assuring rotation invariance during matching used by the multi-scale approaches have an additional negative impact on accuracy.

The processing time required to match features during each iteration of the SLAM system plays an important role in the selection of the detector and the descriptor. The combination of the FAST detector and either the BRIEF or the ORB descriptor is the fastest. The template based matching was four times slower, yet the processing time is still acceptable. The most accurate combination of the FAST detector and the SIFT descriptor was four times slower than the template matching, rendering its usefulness doubtful.

Table 3. The average matching time for various feature point detector and descriptor combinations (in seconds)

Det. \ Descr.	NCC	SURF	ORB	BRIEF	BRISK	FREAK	SIFT
BRISK	0.014	0.658	0.648	0.622	0.638	0.716	0.700
FAST	0.014	0.018	0.004	0.004	0.617	0.083	0.056
GFTT	0.014	0.014	0.022	0.014	0.636	0.102	0.055
Harris	0.012	0.024	0.019	0.021	0.630	0.102	0.048
ORB	0.012	0.381	0.020	0.017	0.641	0.098	0.511
SIFT	0.016	0.167	-	0.164	0.723	0.240	0.159
STAR	0.012	0.031	0.021	0.020	0.633	0.100	0.118
SURF	0.013	0.237	0.144	0.137	0.776	0.227	0.788

5 Conclusions

In this paper, the impact of the selected image feature detectors and descriptors on the visual SLAM was evaluated. The evaluation was performed using real-world data - a test sequence of images acquired by a mobile robot operating in an indoor environment. The availability of accurate ground truth data enabled a direct comparison of the robot trajectory reconstructed using EKF-based SLAM with the reference trajectory, providing informative results on the impact of the choice of the detector-descriptor pair on SLAM accuracy.

In the case of SLAM using single-scale feature detectors and descriptors results in better accuracy and keeps the processing time short. The FASTBRIEF combination was the fastest and the second best in terms of accuracy and can be suggested as the method of choice for indoor visual SLAM. The multi-scale approaches are certainly better suited for tasks where wide baseline matching is necessary. Moreover, the rotational invariance provided by some of the approaches might be beneficial in outdoor scenarios.

Planned future work will be focused on broadening the range of test sequences registered both in indoor and outdoor application scenarios.

Acknowledgment. This research was financed by the Polish National Science Centre grant funded according to the decision DEC-2011/01/N/ST7/05940, which is gratefully acknowledged.

References

1. Agrawal, M., Konolige, K., Blas, M.R.: CenSurE: Center surround extremas for realtime feature detection and matching. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 102–115. Springer, Heidelberg (2008)
2. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: Fast Retina Keypoint. In: Proc. of Int. Conf. on Computer Vision and Pattern Recognition, pp. 510–517 (2012)
3. Aans, H., Dahl, A.L., Pedersen, K.S.: Interesting Interest Points - A Comparative Study of Interest Point Performance on a Unique Data Set. International Journal of Computer Vision 97, 18–35 (2011)
4. Ballesta, M., Gil, A., Martnez, O., Reinoso, M.O.: Local Descriptors for Visual SLAM. In: Proc. Workshop on Robotics and Mathematics (2007)
5. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding 110(3), 346–359 (2008)
6. Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P.: BRIEF: Computing a Local Binary Descriptor Very Fast. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(7), 1281–1298 (2012)
7. Dahl, A.L., Aans, H., Pedersen, K.S.: Finding the Best Feature Detector Descriptor Combination. In: Proc. of Int. Conf. on 3D Imaging, Modeling, Processing, Visualization and Transmission, pp. 318–325 (2011)
8. Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: MonoSLAM: Real-Time Single Camera SLAM. IEEE Transactions on PAMI 29(6), 1052–1067 (2007)

9. Fraundorfer, F., Scaramuzza, D.: Visual Odometry: Part II - Matching, Robustness, and Applications. IEEE Robotics and Automation Magazine 19(2) (2012)
10. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference, pp. 147–151 (1988)
11. Klippenstein, J., Zhang, H.: Quantitative Evaluation of Feature Extractors for Visual SLAM. In: Proc. of the 4 th Canadian Conf. on Comp. and Robot Vision, pp. 157–164 (2007)
12. Leutenegger, S., Chli, M., Siegwart, R.: BRISK: Binary Robust Invariant Scalable Keypoints. In: Proc. of Int. Conv. on Computer Vision, pp. 2548–2555 (2011)
13. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
14. Mozos, Ó.M., Gil, A., Ballesta, M., Reinoso, O.: Interest Point Detectors for Visual SLAM. In: Borrajo, D., Castillo, L., Corchado, J.M. (eds.) CAEPIA 2007. LNCS (LNAI), vol. 4788, pp. 170–179. Springer, Heidelberg (2007)
15. Rosten, E., Drummond, T.W.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
16. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.R.: ORB: An efficient alternative to SIFT or SURF. In: Proc. of Int. Conv. on Computer Vision, pp. 2564–2571 (2011)
17. Schmidt, A., Kraft, M., Kasiński, A.: An evaluation of image feature detectors and descriptors for robot navigation. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part II. LNCS, vol. 6375, pp. 251–259. Springer, Heidelberg (2010)
18. Shi, J., Tomasi, C.: Good Features to Track. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 593–600 (1994)
19. Sturm, J., Magnenat, S., Engelhard, N., Pomerleau, F., Colas, F., Burgard, W., Cremers, D., Siegwart, R.: Towards a benchmark for RGB-D SLAM evaluation. In: Robotics: Science and Systems Conference (2011)

Face Recognition: Shape versus Texture

Maciej Smiatacz

Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland
slowhand@eti.pg.gda.pl

Abstract. This paper describes experiments related to the application of well-known techniques of the texture feature extraction (Local Binary Patterns and Gabor filtering) to the problem of automatic face verification. Results of the tests show that simple image normalization strategy based on the eye center detection and a regular grid of fiducial points outperforms the more complicated approach, employing active models that are able to accurately locate several landmarks. On the other hand, the proposed shape descriptor provides promising results, while the texture features appear to be very sensitive to realistic illumination changes.

1 Introduction

Automatic face recognition has been a popular research area since the early 1990's. It gained a lot of attention for being the iconic symbol of computer vision, the most natural kind of biometrics, and a potential source of many business opportunities related to applications ranging from entertainment to forensics. Although new biometric techniques such as knuckle [1] analysis have emerged, several research papers devoted to face recognition are published every month. The motivation behind many of them is to present yet another mathematical transformation in the context of the leading edge biometric technology. Since real-life tests and comparisons of face recognition methods are complicated and time consuming, standard evaluation protocols do not exist, and reference databases contain photographs taken in laboratories, it is relatively easy to propose solutions that appear to be novel, original and effective. This leads to the pathological situation, when misconceptions and errors propagate and form the foundation for further research.

To address this issues we have developed the Classification Framework (CF) [2] thanks to which many pattern recognition and image processing algorithms, implemented in C++, may be combined into one fully functional face identification system and then tested thoroughly on large data sets. In our research we usually concentrate on the problem of face *verification*, also called the 1 to 1 scenario, which means that the user claims some identity, and the system must decide whether the submitted image is similar enough to the biometric template obtained during the training stage. This task is often more difficult than the typical multiclass *recognition*, which reduces to selecting the most similar template. However, most real-life biometric applications involving user authentication require effective verification, not recognition, procedures.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_26

In our recent work [3] we have shown that the popular linear transformations based on eigenvectors of different kinds of scatter matrices are hardly useful in the context of face verification. In this paper we are trying to answer another set of fundamental questions instead of promoting a new algorithm. The goal of our research was to check whether: 1) the well-known and widely used texture descriptors such as Local Binary Patterns (LBP) and Gabor filters can in actual fact form a robust and illumination invariant set of facial features, 2) the precise feature localization achieved with the help of Active Shape Models increases the stability of the texture description, 3) the shape information alone can be used to perform face verification. In the subsequent sections we will present the methods involved in our tests, the results of experiments, and the conclusions that may be of interest for practitioners working in the field of face recognition.

2 Analyzed Methods

The common intuition among non-professionals is that automatic face recognition relies on some set of geometrical measurements (the distance between the eyes, size of the mouth etc.). However, in the very early years of face recognition this basic concept was discarded, as it appeared [4] that image templates (patches of original photographs) described faces more effectively. Consequently, holistic methods such as Eigenfaces [5] became popular, together with texture-based algorithms using Gabor wavelets [6] or Local Binary Patterns [7]. At the same time, however, automatic shape description became more feasible with the introduction of Active Shape Models [8].

2.1 Local Binary Patterns

Local Binary Patterns were proposed in [7] as a means of describing the gray-level distribution of an image. The LBP operator considers a small (3×3 in the basic case) neighborhood of each pixel, and performs the binarization in this area, using the intensity of the central point as the threshold value. The original form (taking the values of 8 neighbors into account) was defined as:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} 2^n s(i_n - i_c) \tag{1}$$

where (x_c, y_c) denotes the coordinates of the central point, i_c and i_n represent, respectively, the intensity of the central pixel and the intensity of its n-th neighbor, while the $s(u)$ function returns 1 when $u \geq 0$ or 0 otherwise. An image can be described by one global histogram of LBP codes or a set of histograms, created for particular regions; they may be either uniformly distributed (in a form of $n \times m$ grid) or located around some characteristic points (e.g. eye centers, mouth corners and so on). The so-called uniform codes, i.e. binary sequences containing at most one 0-1 and at most one 1-0 transition, are particularly significant. They correspond to important structures such as edges or corners, while

the rest mostly represents the noise. Typically, the LBP histogram contains 58 bins for uniform codes and one for all non-uniform ones. Having computed a set of histograms for two images A and B, we can measure their similarity using the following formula:

$$\chi^2_w(A,B) = \sum_{i,j} w_j \frac{(a_{j,i} - b_{j,i})^2}{a_{j,i} + b_{j,i}} \tag{2}$$

where $a_{j,i}$ denotes the value of the i-th bin in the j-th histogram describing image A, and w_j is the weight representing the importance of j-th region.

The authors of face recognition systems that rely on LBP codes as facial features usually justify their choice by indicating the invariance of the method to the lighting conditions. In fact, the invariance is only guaranteed for monotonic illumination changes, rarely encountered in real life.

2.2 Gabor Filters

Gabor filter is the most widely used advanced method of texture description. It gained particular interest when the biological studies showed that it can be treated as the model of a simple cell in the visual cortex of mammalian brain. It has been used in numerous face recognition algorithms since the pioneering work from 1993 [6] in which the concept of Gabor kernel was introduced:

$$\psi_{\mathbf{k}}(\mathbf{p}) = \frac{\|\mathbf{k}\|^2}{\sigma^2} e^{-\frac{\|\mathbf{k}\|^2\|\mathbf{p}\|^2}{2\sigma^2}} \left(e^{i\mathbf{k}\cdot\mathbf{p}} - e^{-\frac{\sigma^2}{2}} \right) \tag{3}$$

where $\mathbf{p}$ denotes the point for which the value of (3) is calculated and $\mathbf{k}$ is the vector that encodes the wavelength together with the orientation of the filter. Typically, 5 frequencies and 8 orientations are taken into account, resulting in 40-element filter bank. If we consider some neighborhood S of image I around the point $\mathbf{p}$, then the following convolution:

$$W_I(\mathbf{k},\mathbf{p}) = (\psi_k * I)(\mathbf{p}) = \sum_{\mathbf{p}_i \in S} \psi_k(\mathbf{p} - \mathbf{p}_i) I(\mathbf{p}_i) \tag{4}$$

will provide the local texture description in the form of the so-called *jet*, i.e. the 40-element vector $\mathbf{j}$ containing complex numbers j_n. Usually only the magnitudes m_n of j_n coefficients are further processed, as they vary slowly with the position. According to [6], it is advised to use the cosine metrics to compare two jets $\mathbf{j}$ and $\mathbf{j'}$:

$$S_m(\mathbf{j},\mathbf{j}') = \frac{\sum_{i=1}^{40} m_i m_i'}{\sqrt{\sum_{i=1}^{40} m_i^2 \sum_{i=1}^{40} m_i'^2}} \tag{5}$$

as it proved to be robust with respect to the global changes in contrast induced by varying illumination.

2.3 Active Shape Models

Active Shape Models (ASM) were introduced in the early 1990's [8] as a technique allowing for precise localization of objects in images. The model of this kind is based on a template of the particular object, represented by a set of characteristic points (landmarks). During the training stage the model retrieves the information from the (usually large) set of manually labeled examples. It is able to learn the mean shape of the object, $\bar{\mathbf{\Gamma}}$, as well as the main modes of variation (most important deformations) which can be observed in the training set. The latter are stored as the columns of the transformation matrix $\mathbf{P}$. The new instance of the model, $\mathbf{x}$, can be obtained by changing the vector of parameters $\mathbf{b}$:

$$\mathbf{x} = \bar{\mathbf{\Gamma}} + \mathbf{P}\mathbf{b} \tag{6}$$

On the other hand, $\mathbf{b}$ can be viewed as a feature vector describing the shape $\mathbf{x}$.

The algorithm that fits the model to the image, apart from modifying its position, orientation and scale, is also able to change the model's shape, but only within some limits preventing unnatural deformations. The ASM evolved into several kinds of Active Appearance Models (AAM) [9], in which the texture information is merged with data describing the shape. This combination usually increases the accuracy of the fitting process.

Intuitively, the application of an active deformable model should be a perfect way to steer the extraction of texture features: once the landmarks are precisely localized, image fragments representing the most important regions of a face may be described in terms of the significant structures that exist around them. We decided, however, to experimentally check the correctness of this notion. Additionally, we realized that the model fitted properly to the image can be treated as a source of information about the shape of the face. Therefore we propose to use the coordinates of the landmarks directly to form a kind of a chain code $\mathbf{f}$, a vector containing the following elements:

$$f_0 = x_0, f_1 = y_0, ..., f_{2*i} = (x_i - x_{i-1}), f_{2*i+1} = (y_i - y_{i-1}) \qquad i = 1..N \tag{7}$$

where N is the number of landmarks, and use it as a simple face shape descriptor.

3 Experimental Face Verification System

The experiments were carried out with the help of our Classification Framework that allowed us to compose several variants of the face verification system from the individual components implemented as C++ classes. Fig. 1 shows the architecture of our experimental application.

In order to test the importance of precise feature localization two kinds of image normalizers were used. The first one (called AB) was based on the well-known approach described in [10], i.e. the object detector combining adaptive boosting with hierarchical classification. This method determines positions of eye centers, rotates the face to the upright position, and measures the interocular distance d_e. Next, using the value of d_e, the face region is cropped (Fig. 2a) and

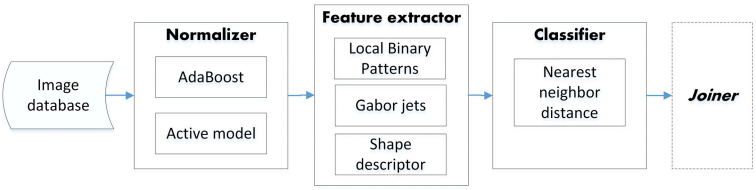


Fig. 1. Main building blocks of the experimental system

resized to the rectangular shape (64×64 pixels). The second normalizer (ASM) employs an active model to determine the precise localization of the landmarks. Although we developed our own implementation of Active Appearance Models [11], it is still not ready for the batch processing of large data sets. Therefore, we decided to rely on the commercial face localization web service provided by Betaface [12]. After uploading the image, encoded with Base64 algorithm as a part of XML-formatted message, the service returns coordinates of the 94 characteristic points, together with other parameters, such as the angle that can be used to rotate the face to the vertical position. Using the landmark points we calculate the bounding box of the face and then increase its size by 10% along both axes.

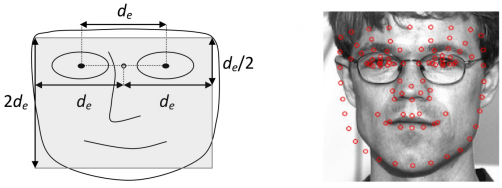


Fig. 2. (a) Image normalization based on eye center detection, (b) typical ASM localization error

Feature extractor contains implementations of Local Binary Patterns (1), Gabor filters (4) and our simple shape descriptors (7). In order to compare just the different types of features we refrain from using any further transformations, and resort to the nearest neighbor classifier. For each person P a set of M_P training images t_i is collected, and since we are interested in face verification, a test sample x is accepted as representing the legitimate user if

$$\exists \mathbf{t}_i \quad D(\mathbf{x}, \mathbf{t}_i) < \alpha \cdot w \qquad i = 1..M_P \tag{8}$$

where $D()$ is the distance defined by (2) for LBP features, (5) for Gabor filtering, and by Euclidean distance for shape descriptors; α is the maximum distance between training samples t_i, and w is a tuning parameter that controls the balance between the false acceptance and false rejection rates.

Classification Framework also offers the possibility of using several parallel pipelines that independently process input images. For example, one image may be described by two different feature vectors, requiring different kinds of similarity measures. In such a case we apply the optional joiner, which calculates the final value of D from the distances D_j returned by individual pipelines, using the *z-score* method:

$$D(\mathbf{x}) = \sum_j \left(D_j(\mathbf{x}) - \mu_j\right)/\sigma_j \tag{9}$$

where μ_j and σ_j, are, respectively, the mean and standard deviation of D_j over the training set.

4 Results

The photographs from the CMU-PIE database [13] were used throughout the experiments. For each of the 68 persons represented in the data set, the images taken from two almost frontal viewpoints (cameras no 7 and 27) during the expression session (showing the subjects with a neutral expression, smiling, and blinking) were selected as the training set (from 2 to 8 samples per person, 6 in most of the cases). The testing set was divided into two parts: F, including about 1200 nearly frontal images (cameras 7, 9, 27) taken under the same neutral illumination conditions, and L containing also around 1200 frontal images (taken by camera 27 only), but with flash fired from several different positions. The F set was supposed to check the robustness of the features to the small changes of pose and expression, while L set was prepared to test the illumination invariance of texture descriptors. The training and testing procedure was repeated 68 times for every verification method – Table 1 summarizes the results. The average FAR and FRR coefficients supplement the rates calculated separately for F and L sets. After preliminary experiments different values of weight w (8) were assigned to the feature extraction methods. One configuration including two pipelines was tested (LBP features calculated from the regular grid plus the chain code (7) extracted from active model), together with the simplest reference solution based on pixel-by-pixel comparisons of 60×40 templates extracted from eyes and nose regions.

The analysis of our experimental results leads to the following observations. 1) Local Binary Patterns are very sensitive to realistic (not monotonic) illumination changes (the FRR is dramatically higher for the L data set). Gabor filters show some level of illumination invariance, but at the same time they are less discriminant than LBPs (FAR is worse) and about 50 times more computationally expensive. 2) Precise localization of landmarks does NOT improve the system's performance; on the contrary, application of the active model radically increased the error rates, especially in the case of Gabor filters. This is surprising only at the first sight. Although the AB normalizer uses the regular grid, it also locates the eye centers at the same position for every face image, so that the texture parameters extracted at the nodes of the grid also encode the shape variations to some extent. On the other hand, precise feature localization corresponds to

Table 1. The results of experiments (FAR and FRR values given in percent)

Normalizer	Features	Weight	Set	FAR	FRR	Avg. FAR	Avg. FRR
AB	LBP	0.75	F	0.17	23.15	0.09	57.34
			L	0.01	91.53		
	Gabor	0.5	F	1.84	22.81	1.14	28.21
			L	0.44	33.61		
ASM	LBP	0.75	F	10.58	13.54	8.59	33.12
			L	6.59	52.7		
	Gabor	0.5	F	9.39	70.93	9.86	74.77
			L	10.32	78.6		
	Shape	0.25	F	5.34	39.73	4.84	39.66
			L	4.33	39.58		
ASM+AB	LBP + Shape	0.001	F	7.1	22.27	4.46	42.25
			L	1.82	62.22		
ASM	Eyes & nose	0.75	F	1.3	17.81	0.80	38.71
			L	0.3	59.6		

looking closely at the same points of different faces. In such conditions every eye corner or nose tip becomes similar to the others, so it is more difficult to distinguish between faces. 3) The results achieved by the variant that used the shape information only are quite promising, especially if we take into account that the descriptor (7) is very simple. Moreover, even the state-of-the-art implementations of active models are still much less accurate than we might expect, mainly when it comes to marking the outer boundary of a face. As a result many distinctive face features, such as the characteristic shape of the chin (Fig. 2b) are usually lost. As expected, this approach is completely illumination invariant. On the other hand, combining the texture features with shape description using the z-score (8) does not lead to any improvements. 4) The simplest algorithm relying on eye and nose templates performed unexpectedly well, even under varying lighting conditions.

5 Conclusion

Our experiments show that, contrary to the popular belief, the accurate localization of landmark points only worsens the performance of a face verification system based on the texture features. Additionally, neither Local Binary Patterns nor Gabor filtering provide satisfactory level of illumination invariance. Although after 20 years the recommendations formulated in [4] still remain in force (templates perform better than geometrical parameters), in our opinion the future of automatic face identification is not related to the texture descriptors, but the improvements of active models that would finally allow us to extract the precise shape information from face images.

References

1. Choras, M., Kozik, R.: Contactless palmprint and knuckle biometrics for mobile devices. Pattern Analysis and Applications 15, 73–85 (2012)
2. Smiatacz, M., Przybycien, K.: A Framework for Training and Testing of Complex Pattern Recognition Systems. In: IEEE Conf. on Signal and Image Processing Applications (ICSIPA) (2011)
3. Smiatacz, M.: Eigenfaces, Fisherfaces, Laplacianfaces, Marginfaces – How to Face the Face Verification Task. In: Burduk, R., Jackowski, K., Kurzynski, M., Wozniak, M., Zolnierek, A. (eds.) CORES 2013. AISC, vol. 226, pp. 191–200. Springer, Heidelberg (2013)
4. Brunelli, R., Poggio, T.: Face recognition: features versus templates. IEEE Transactions on PAMI 15, 1042–1052 (1993)
5. Turk, M., Pentland, A.: Eigenfaces for Recognition. J. Cognitive Neuroscience 3(1), 71–86 (1990)
6. Lades, M., Vorbruggen, J.C.: Distortion invariant object recognition in the dynamic link architecture. IEEE Computers 42, 300–310 (1993)
7. Ojala, T., et al.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 29, 51–59 (1996)
8. Cootes, T.F., et al.: Active shape models – their training and application. Comp. Vision and Image Understanding 61, 38–59 (1995)
9. Cootes, T.F., et al.: Active Appearance Models. IEEE Trans. PAMI 23, 681–685 (2001)
10. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. of Computer Vision 57, 137–154 (2004)
11. Smiatacz, M., Sikora, D.: AAM Toolkit: a system for visual object appearance modeling. In: Nguyen, N.T., Zgrzywa, A., Czyżewski, A. (eds.) Advances in Multimedia and Network Information System Technologies. AISC, vol. 80, pp. 121–129. Springer, Heidelberg (2010)
12. `http://www.betafaceapi.com/`
13. Phillips, P.J., et al.: The FERET evaluation methodology for face recognition algorithms. IEEE Trans. PAMI 22, 1090–1104 (2000)

Spatio-temporal Digital Path Approach to Video Enhancement

Marek Szczepański

Faculty of Automatic Control, Electronics and Computer Science,
Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
Marek.Szczepanski@polsl.pl

Abstract. An overview of the new fast and efficient spatio-temporal video filtering technique was presented in this paper. The new approach is based on digital paths concepts in three dimensional space. The digital paths can explore image structures in spatial as well as temporal coordinates from subsequent frames. Presented technique copes with different video artifacts such as Gaussian, impulsive and grain noise and still preserves and even enhances edges. The new method can even remove JPEG artifacts and video flickering. Preliminary results show that the proposed algorithm can be used both for offline and online processing.

1 Introduction

Imaging sensors used in modern cameras or industrial cameras are getting better and closer to the theoretical limits of their physical capabilities. However, they are still commonly used sensors with much lower specs, they are used in web cameras, cell phones and monitoring systems. The sensor noise can be a serious problem even in high-end the high-speed cameras especially when capturing videos in low light environment. Is therefore necessary to use algorithms that improve the quality of such images and sequences. One of the most challenging tasks is shot noise removal which is dominant in low light images [5].

Several noise suppression 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 [1,13,6,10]. However, standard image processing techniques cannot utilize all available information i.e. similarities in neighboring frames. 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. One of the simplest temporal filters is the Temporal Gaussian Filter (*TGauss*). Simple temporal filters are fast and very effective for static sequences, however averaging frames where there are objects in motion creates "ghosting"effects in output video.

Some algorithms utilize motion compensation to reduce blurring effect such as works presented by Dubois and Sabri [4] and later in work [14]. Another solution

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_27

to reduce ghosting artifacts is to use a temporal version of bilateral filter used as a element of ASTA filter [2].

In this paper we propose a different approach. We propose to exploit possible connections between successive image pixels using the concept of digital paths in spatial domain which can be extended to temporal domain and understood as trajectories or object displacements in subsequent frames.

Proposed algorithm is based on concept of two-dimensional Digital Path Approach presented in papers [10]. According to the proposed here methodology, image pixels are grouped together forming paths that reveal the underlying structural dynamics of the image. Although the proposed algorithm does not apply motion compensation, a similar effect is obtained by using paths that track moving objects in a sequence of images

The proposed filtering technique can successfully eliminate Gaussian and impulsive noise as well as digital compression artifacts and even flickering in input video sequences.

The paper is organized as follows. In Section 2 the general concept of the digital paths approach is introduced. while Section 3 presents simulation results. Finally, Section 4 summarizes our paper.

2 Digital Paths Approach

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 spatio-temporal window W:

$$\hat{\mathbf{F}}_{\mathbf{0}} = \sum_{i=0}^{k-1} w_i \mathbf{F}_i = \frac{\sum\limits_{i=0}^{k-1} \mu_i \mathbf{F}_i}{\sum\limits_{i=0}^{k-1} \mu_i}, \tag{1}$$

where $\mathbf{F}_i$ and $\hat{\mathbf{F}}_{\mathbf{0}}$ denotes filter inputs and output respectively.

In our case weights will be calculated using similarity functions calculated over digital paths included in processing window W in spatio-temporal domain.

Let a digital lattice $\mathcal{H} = (\mathbf{F}, \mathcal{N})$ be defined by $\mathbf{F}$, which is the set of all points of the processing window W in spatio-temporal domain and a neighborhood relation $\mathcal{N}$ between the lattice points [8].

A digital path $P = \{p_i\}_{i=0}^{n}$ defined on the lattice $\mathcal{H}$ is a sequence of neighboring points $(p_{i-1}, p_i) \in \mathcal{N}$. The length $L(P)$ of the digital path $P\,\{p_i\}_{i=0}^{n}$ is simply $\sum_{i=1}^{n} \rho^{\mathcal{H}}(p_{i-1}, p_i)$, where $\rho^{\mathcal{H}}$ denotes the distance between two neighboring points of the lattice $\mathcal{H}$. In the case of a three-dimensional spatio-temporal window three basic neighborhood system can be defined $\mathcal{N}_{26}$, $\mathcal{N}_{18}$ and $\mathcal{N}_6$ (Fig. 1). An $\mathcal{N}_{26}$ and $\mathcal{N}_{18}$-neighborhood systems are considered in this work with a topological distance 1 assigned between two neighboring points.

Let us adopt the following notation, which will help us define the distance functions defined over digital paths. Successive path points successive path points

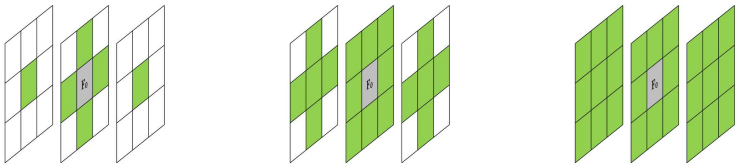


Fig. 1. Spatio-temporal masks for different neighborhood systems

denoted as will be denoted as $p_0 = (x_0, y_0)$, $p_1 = (x_{u_1}, y_{v_1})$,$p_2 = (x_{u_2}, y_{v_2})$ and so on, till the path reaches in n steps the ending point $p_n = (x_{u_n}, y_{v_n})$. The set of all possible digital paths contained in W joining two points $a, b \in W$ will be denoted as $\Phi^W(a, b)$. Two pixels a and b will be called connected (hereafter denoted as $a \leftrightarrow b$), if there exists a digital path $P^W(a, b)$ contained in the set W starting from a and ending at b.

If two pixels p_0 and p_n are connected by a digital path $P^{W,n}\{p_0, p_1, \ldots, p_n\}$ of length n then let $\Lambda^{W,n}\{p_0, p_1, \ldots, p_n\}$ be a function which measures the connection cost defined over the digital path linking the starting point p_0 and ending point p_n. The connection cost over the digital path $\Lambda^{W,n}$ will be defined as a measure of dissimilarity between color image pixels p_0, p_1, ..., p_n forming a specific path linking p_0 and p_n [3,12]:

$$\begin{aligned} &\Lambda^{W,n}\{p_0, p_1, p_2, \ldots, p_n\} \\ &= f\{\mathbf{F}(p_0), \mathbf{F}(p_1), \mathbf{F}(p_2), \ldots, \mathbf{F}(p_n)\} = \sum_{i=1}^{n} \|\mathbf{F}(p_i) - \mathbf{F}(p_{i-1})\|. \end{aligned} \tag{2}$$

Let us now define a similarity function between the starting point $a = p_0$ and point $b = p_1$ crossed by the digital path connecting pixel p_0, its neighbor p_1 with all possible points p_n which can be reached in n steps from p_0.

The aim of taking into account the points p_2, ..., p_n when calculating the similarity between p_0 and p_1 is to explore not only the direct neighborhood of p_0 but also to use the information on the local image structure. This can be done by acquiring the information on the local image features investigating the connection costs of digital paths originating at p_0, passing p_1 and then visiting successive points, till the path reaches length n. This approach will be further denoted as $DPA_{1st}3D$ to distinguish it from two-dimensional DPA_{last} algorithm proposed in the work [10]. In this case the similarity function takes the form:

$$\mu^{W,n}(a, b) = \mu^{W,n}(p_0, p_1) = \sum_{m=1}^{\omega} g\left(\Lambda_m^{W,n}(p_0, p_1)\right), \tag{3}$$

where ω denotes number of all possible paths $P\{p_0, p_1, p_2^*, \ldots, p_n^*\}$ of length n originating at $a = p_0$ and crossing $b = p_1$ totally included in processing window W, $\Lambda_m^{W,n}\{\cdot\}$ is a dissimilarity value along a specific path and $g(\cdot)$ is a smooth function of $\Lambda_m^{W,n}$.

In this work we assume that $g(\cdot)$ is the exponential function [10] so our similarity function takes the form:

$$\mu^{W,n}(a,b) = \mu^{W,n}(p_0,p_1) = \sum_{m=1}^{\omega} \exp\left[-\beta \cdot \Lambda_m^{W,n}(p_0,p_1)\right] \tag{4}$$

where β is the filter design parameter. A normalized form of the similarity function can be defined as follows:

$$\psi^{W,n}(a,b) = \psi^{W,n}(p_0,p_1) = \frac{\mu^{W,n}(p_0,p_1)}{\sum\limits_{p_1^*} \mu^{W,n}(p_0,p_1^*)}, \tag{5}$$

where p_1^* denotes all p_0 neighbors.

Assuming that the pixel $a = p_0$ is the pixel under consideration,with $\mathbf{F}(b)$ representing the pixel $b = p_1$ the filter output $\hat{\mathbf{F}}(a)$ is given as follows:

$$\hat{\mathbf{F}}(a) = \sum_{b \sim \mathcal{N}_{26}(a)} \psi^{W,n}(a,b) \cdot \mathbf{F}(b) = \sum_{p_1^*} \psi^{W,n}(p_0,p_1^*) \cdot \mathbf{F}(p_1^*). \tag{6}$$

The performance of the new filters strongly depends on the type of digital paths selected. Different models of paths result to application-specific filters, which are able to suppress certain types of noise [10]. In this paper we concentrate on the *"Self-Avoiding Path model"* (SAP), which provides a model suitable for image processing applications [9,10]. In the *Self-Avoiding Path* (SAP) no vertex is visited more than once resulting in a trajectory that never intersects itself.

The computational complexity of the proposed filter depends mainly on the number of paths formed, we want therefore to minimize the quantity of created paths without losing the effectiveness of the filter. All further discussion will concern paths of length $n = 2$, as it has been shown that they give satisfactory filtering results at a relatively low computational cost [10]. Table 1 shows the dependence of the number of possible paths to form on the spatio-temporal window size. In order to reduce filter complexity the size of the supporting window W can be limited to $(3 \times 3 \times 3)$ independently of the path's length, so we obtain 168 possible paths for $\mathcal{N}_{18}$ neighborhood system. Such filter will be denoted as $FDPA_{1st}3D$ (Fast Digital Paths Approach). Additionally calculation of path lengths can be parallelized using the GPU. Preliminary results wit OpenCL technology show that we can in this way obtain real-time processing of HD sequences.

Table 1. The number of self-avoiding paths of length two, depending on the 2D window radius A, temporal radius t and the selected neighborhood system

Neighborhood	$\mathcal{N}_6$		$\mathcal{N}_{18}$		$\mathcal{N}_{26}$	
A	1	2	1	2	1	2
$t = 1$	24	28	**168**	256	**264**	488
$t = 2$	26	30	210	306	362	650

3 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 *PSNR* and the *SSIM* [15] were used for the analysis. The new filter performance was compared to the following methods: 1) Wiener2D - a spatially adaptive Wiener filter, 2) Wiener 3D -our implementation of a simple extension of Wiener2 to three dimensions, with a window size of $3 \times 3 \times 3$, 3) temporal Gaussian Filter *TGauss* (with time window $n = 5$ and $\sigma = 5$), 4) Vector Median Filter (with window 3×3 and L_1 norm)[1], 5) Spatial Fast Digital Paths Approach *FDPA* (L_1 norm, $\beta = 15$) [10], 6) Spatio-Temporal Fuzzy FDPA Filter (*STFFDPA*) (L_1 norm, $n = 5$, $\sigma = 5$, $\gamma = 4$, $\beta = 15$)[11]. Parameters of the new filters are set empirically to match the wide range of sequences and different noise scenarios so we used $\mathcal{N}_{18}$and $\mathcal{N}_{26}$-neighborhood, L_1 norm and $\beta = 15$. For all digital paths based filters, the path length was limited to two steps.

Fig. 2 shows the frame from noisy sequence captured in low light conditions containing small toy-car moving rapidly and Fig. 3 depicts filtering results of sample frame from standard video sequence *Hall Monitor* corrupted with Gaussian noise ($\sigma = 10$). It can bee 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.

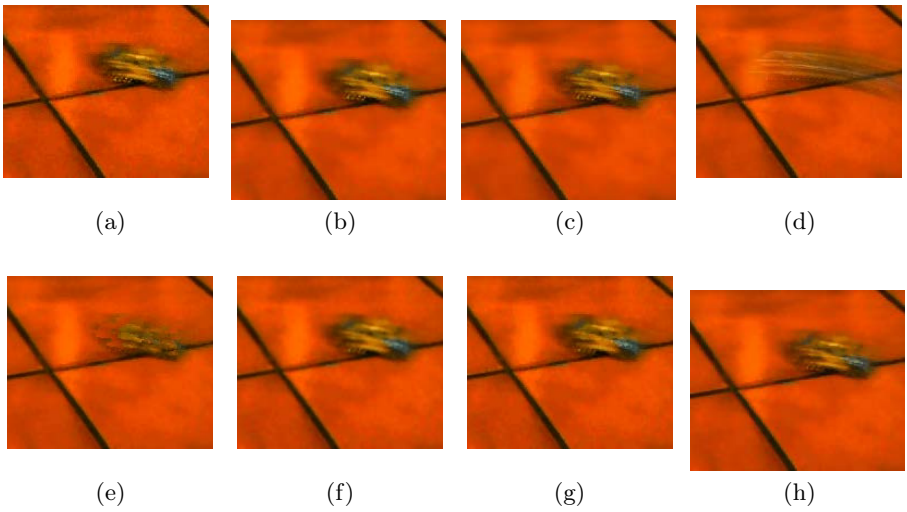
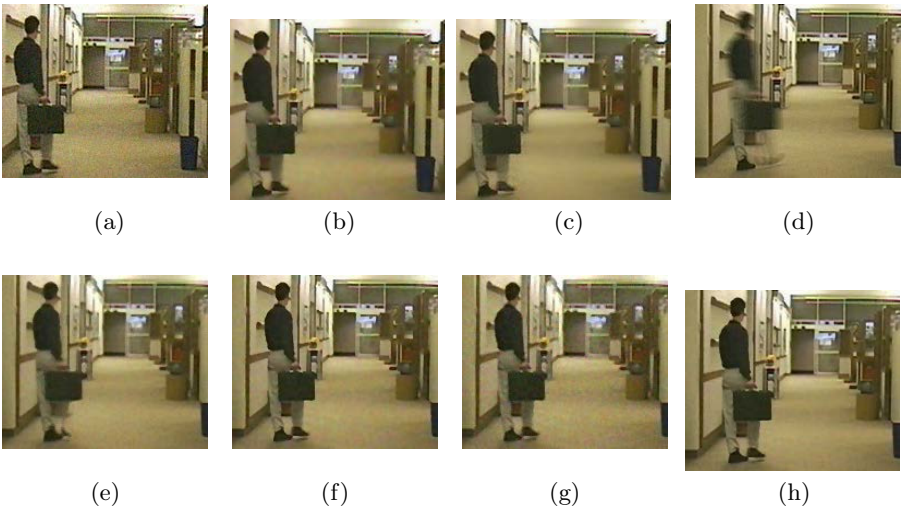


Fig. 2. (a) Frame from the noisy video seqence - *Toy Car* and results of filtering with (b) Wiener2D,(c) Wiener3D, (d) Temporal Gaussian, (e) *VMF3D*, (f) Spatial *FDPA*, (g) *VMF* and (h)New Spatio-Temporal filter - $FDPA_{1st}3D$



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 3. (a) Frame from the noisy video sequence *Hall Monitor* with Gaussian noise ($\sigma = 10$) and results of filtering with (b) Wiener2D,(c) Wiener3D, (d) Temporal Gaussian, (e) *VMF3D*, (f) Spatial *FDPA*, (g) *VMF* and (h) New Spatio-Temporal filter - *FDPA*$_{1st}$*3D*

Table 2. Comparison of the filtering algorithms applied for two standard test sequences corrupted with different noise scenarios

Video Sequence	*Foreman*				*Hall Monitor*			
Noise Scenario	G_{10}	G_{20}	GI_{10}	GI_{20}	G_{10}	G_{20}	GI_{10}	GI_{20}
	PSNR Results [dB]							
Noisy	28.14	22.22	17.75	14.43	28.25	22.29	17.91	14.57
Wiener2D	32.83	27.83	21.54	18.91	32.34	27.69	21.79	19.01
Wiener3D	32.05	27.91	22.64	19.61	32.33	28.05	22.87	19.71
VMF	29.57	25.78	29.16	24.85	28.78	25.48	28.37	24.51
VMF3D	28.71	26.65	28.54	26.16	29.44	27.35	29.28	26.84
FDPA	31.90	28.51	30.24	24.36	31.66	28.10	29.34	24.08
TGauss	27.89	25.72	22.88	20.21	32.39	28.25	24.38	21.08
STFFDPA	31.59	29.65	28.93	23.50	33.10	28.53	29.29	23.63
FDPA$_{1st}$**3D**$\mathcal{N}_{18}$	**33.64**	30.19	**31.08**	26.28	**34.56**	**30.71**	**31.79**	26.97
FDPA$_{1st}$**3D**$\mathcal{N}_{26}$	33.55	**30.25**	30.94	**26.29**	34.25	**30.71**	31.51	**26.99**
	SSIM Results							
Noisy	0.773	0.521	0.347	0.191	0.758	0.513	0.352	0.214
Wiener2D	0.906	0.779	0.493	0.334	0.922	0.796	0.514	0.350
Wiener3D	0.898	0.784	0.529	0.348	0.928	0.809	0.556	0.367
VMF	0.826	0.671	0.814	0.623	0.850	0.688	0.838	0.642
VMF3D	0.815	0.718	0.809	0.693	0.888	0.789	0.883	**0.764**
FDPA	0.894	0.794	0.859	0.643	0.919	0.805	0.877	0.650
TGauss	0.792	0.674	0.450	0.316	0.908	0.768	0.540	0.399
STFFDPA	0.892	0.820	0.840	0.632	0.903	0.775	0.869	0.648
FDPA$_{1st}$**3D**$\mathcal{N}_{18}$	**0.917**	0.840	**0.870**	**0.701**	**0.948**	0.872	**0.917**	0.752
FDPA$_{1st}$**3D**$\mathcal{N}_{26}$	0.916	**0.843**	0.868	0.700	**0.948**	**0.876**	**0.917**	0.757

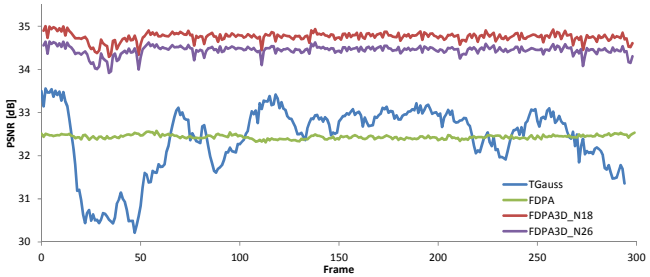


Fig. 4. PSNR coefficients of noisy *Hall monitor* sequence with *TGauss*, *FDPA* and $FDPA_{1st}$ *3D filters*

Average values of objective quality measures for sequence *Foreman* and *Hall Monitor* corrupted with different noise types (Gaussian and mixed Gaussian with Impulsive) are presented in Table 2.

The Fig. 4 depicts PSNR values of subsequent frames filtered with temporal Gaussian, Spatial FDPA and the new spatio-temporal filters, it can be clearly seen when temporal filter produces ghosting artifacts.

4 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. The new approach utilizing digital paths concepts in three dimensional space was presented in this paper. The digital paths can explore image structures in spatial as well as temporal coordinates from subsequent frames. Presented technique provides excellent noise suppression, especially for low light video sequences.

Acknowledgements. This work was supported by the Polish National Science Center (NCN) under the Grant: DEC-2012/05/B/ST6/03428.

References

1. Astola, J., Haavisto, P., Neuovo, Y.: Vector median filters. IEEE Proc. 78, 678–689 (1990)
2. Bennett, E.P., McMillan, L.: Video enhancement using per-pixel virtual exposures. ACM Trans. Graph. 24(3), 845–852 (2005)

3. Cuisenaire, O.: Distance transformations: fast algorithms and applications to medical image processing. PhD thesis, Universite Catholique de Louvain (October 1999)
4. Dubois, E., Sabri, S.: Noise reduction in image sequences using motion-compensated temporal filtering. IEEE Transactions on Communications 32(7), 826–831 (1984)
5. 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)
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.N., Androutsos, D., Venetsanopoulos, A.N.: Fuzzy adaptive filters for multichannel image processing. Signal Processing Journal 55(1), 93–106 (1996)
8. Schmitt, M.: Lecture notes on geodesy and morphological measurements. In: Proceedings of the Summer School on Morphological Image and Signal Processing, Zakopane, Poland, pp. 36–91 (1995)
9. Smolka, B., Wojciechowski, K.: Random walk approach to image enhancement. Signal Processing 81(3), 465–482 (2001)
10. Szczepanski, M., Smolka, B., Plataniotis, K.N., Venetsanopoulos, A.N.: On the geodesic paths approach to color image filtering. Signal Processing 83(6), 1309–1342 (2003)
11. Szczepański, M.: Spatio-temporal fuzzy fdpa filter. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011, Part II. LNCS, vol. 6855, pp. 316–323. Springer, Heidelberg (2011)
12. Toivanen, P.J.: New geodesic distance transforms for gray scale images. Pattern Recognition Letters 17, 437–450 (1996)
13. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: ICCV, pp. 839–846 (1998)
14. Varghese, G., Wang, Z.: Video denoising based on a spatiotemporal gaussian scale mixture model. IEEE Transactions on Circuits and Systems for Video Technology 20(7), 1032–1040 (2010)
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)

Emotion Recognition for Affect Aware Video Games

Mariusz Szwoch and Wioleta Szwoch

Gdansk University of Technology
{szwoch,wszwoch}@eti.pg.gda.pl

Abstract. In this paper the idea of affect aware video games is presented. A brief review of automatic multimodal affect recognition of facial expressions and emotions is given. The first result of emotions recognition using depth data as well as prototype affect aware video game are presented.

1 Affect Aware Video Games

Affective computing is one of the active research areas on human behavior studies and has a significant impact in many fields such as healthcare, education, entertainment etc. [2,3]. Affective computing studies how to automatically recognize, interpret, and process human emotions via analyzing available sensory data. Affect-aware applications, as their additional functionality, react to user emotions.

Video games belong to the wide area of entertainment applications. Though, video games are among some of the most natural applications of affect only few of them seek to incorporate their players affective state into the gameplay. Such games can be referred as affective or more properly affect-aware games. Unfortunately, this affect-awareness is usually statically built-in the game at its development stage basing on the assumed model of so called representative player [4]. There are two problems with such attitude. Firstly, each player differs in some way from that averaged model. Secondly, and more important, players affect state can change even radically from session to session making almost impossible to predict the current user emotions at the development stage. That is why the real-time recognition of players affect may become such important for video games industry in the nearest future.

For the last several years only a few truly affect-aware games have been developed, mainly as noncommercial academic projects. As an example, "Feed The Fish" takes a players facial expressions as input and dynamically responds to the player by changing the game elements [5]. The goal of this system is to use human expressions to build a communication channel between the game and players so playing the game can be more enjoyable [5].

Another attempt is to create a universal architecture for affect-aware games. In [1] a Koko library framework is proposed that abstracts an affect model and sensors handling from other game components which are application logic and

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_28

the game engine. As a proof of concept two sample games had been developed basing on simple physiological and physical signals such as heart rate, skin conductance, or GPS players location.

Nowadays commercial games put a focus on engaging players emotionally through both gameplay techniques and interaction with virtual agents or human players in multiplayer games. But it is quite probable that this situation will change in the nearest future due to development of affect recognition techniques and new input devices and non-intrusive sensors allowing to observe the player in different information channels or modalities. For example, Valve Software, expect affect emotion as essential element of future games, and experimented with biometrics directly by introducing them into a special build of Left 4 Dead 2 [6].

2 Automatic Affect Recognition

Even the best emotional model is able only to predict the current humans affect state, that is why affect recognition is so important in affective software dealing with real users as it try to recognize real not estimated emotions.

There are several affect models developed for different application fields. Some models define continuous emotional space of two or more dimensional coordinate system and place all emotions within this model space. The best known example of this approach is PAD emotional state model defining three dimensions: Pleasure, Arousal, and Dominance, which express respectively, pleasure, intensity and the dominant nature of the emotion [7]. Using this model several different emotion sets may be defined containing 6, 7, or more affect states [8]

Emotion, or affect recognition is a key to create a truly affect-aware video games or other software. It may seem that affect recognition techniques could be applied in applications in any field but, in general, different applications differ in affect model of the user and possible sets of information inputs. For example, health care applications more probably seek for pain or mental illness symptoms allowing to use wired or even intrusive sensors to receive very trusty information. On the other side, affective e-learning and training systems seeks rather for users attention, engagement or boredom and do not demand any other input devices than internet camera and microphone. Affect-aware video games seems similar to e-learning software in the aspect of expected affect states which, however, can be extended by a fear, excitement and other emotions which are not so probable in other applications. Video games, also, may demand some additional controllers and gamers are usually open for new technical solutions that could improve games playability.

So, there are two important assumptions of affect recognition model that should be defined for specific application. The primary is the affect model and the subset of emotions to be recognized. The other is the definition of input devices and sensors to be used.

The goal of automatic emotion recognition is to recognize the current affect state of the user basing on potentially multimodal input signals about the human appearance or physiological parameters as well as other information such as

environmental conditions, application mode, the interaction history, etc. The research in this field has been performed since early 2000., attracting the interest of artificial intelligence and computer vision research communities. Most attempts use face as a primary (and most often the only) source of information trying to recognize facial expressions from static images and video recordings [9,10,11]. This attitude proved to be successful in the case of good lighting and exposure conditions. Other research focused on other expressions of human emotions such as speech, gestures, gaze, and others [11]. Also physiological signals are often used especially in health care systems [2].

Numerous research projects proved that multimodal approach gives in general better results, in terms of recognition efficiency, then taking into account single data input [2]. Combination of multiple types of inputs from different modalities or different features over the same modality vastly significantly improves the system classification abilities [12]. In general, there are two main multi-modal information fusion attitudes. In the first attitude, called early, or feature-level fusion, techniques and methods for affect recognition feed a set of features into a commonly used classifier at the feature level. In the other approach, called late, or decision-level fusion, the recognition is based on some kind of decision systems which input is fed by the classification results of monomodal classifiers [1,12]. The decision system can be a voting expert system, decision tree, belief network, and other.

It is also possible to create hybrid systems in which classifiers of the first level can also use multimodal data, e.g. from a single channel like video or physiological signals.

The process of affect recognition does not differ in its essence from other classification systems such as optical character recognition (OCR), face recognition etc. Classification assumes existence of classifier C that decides about belongingness of an object o_i to one of predefined classes x_j. The object being classified is represented by a feature vector $\boldsymbol{y}$ which can contain features of any type extracted from the object (Fig. 1). In most applications a supervised learning is used to create a classifier on the base of labeled set of objects representative for each class x_j. There is a variety of classifiers type from simple to complex ones that can be used in almost each classification task or only in specific applications. Also, many existing classification algorithms may be adapted to new tasks. There are a large number of software libraries that make available many classifiers of different kind. These classifiers should be fed by properly extracted features (Fig. 1). Thus, the crucial matter is to properly select channels to acquire information about the classified objects, and properly preprocess it. Efficient solving of any recognition problem usually requires researchers with knowledge and pragmatic experience in this field. In most cases, the recognition process uses the same general scheme and proceeds through the same stages (Fig. 1). Acquired data usually needs some initial preprocessing (e.g. filtering, noise removal) and optional segmentation (e.g. background removing). From the preprocessed data a set of different features can be extracted. If the number of features is very large only most informative can be selected to form a feature

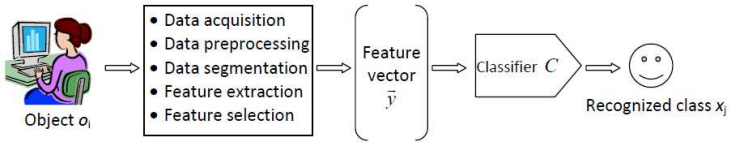


Fig. 1. General recognition process

vector that indirectly characterizes the object. Finally, the classifier C classifies an object into one of possible classes x_j .

2.1 Emotion Recognition from Video Channel

Many affect recognition systems are based on video channel taking into account facial expressions, gestures and posture analysis in mono-modal as well as in multi-modal systems [2,9,10,15]. This situation results from non-intrusive property of cameras and becomes more and more practical with the rapid development of hardware capabilities and computer vision technology. Moreover, vision channel is the most informative, as human beings tend to express their emotions in a visual way.

Automatic analysis of facial expression from video is one of the most common approaches in affective computing [18]. The most extensive work on facial expression analysis is summarized in several survey papers [10,11]. The most advanced algorithms define human face models with characteristic control points and a set of transformations of those points for each recognized facial expression. The active appearance model (AAM) is a statistical approach that models the shape and texture of a target object and has a great success in modeling and recognition of human faces. Traditional AAM framework can fail when the face pose changes as only 2D information is used to model a 3D object. To overcome this limitation, different 3D extensions of AAM are proposed [20] as well as 3D Morphable Models (3DMM) [21]. Both attitudes has a great recognition efficiency of facial expressions and their main limitation is high computing complexity.

Other attitudes use detection of facial characteristic points (e.g. eyes, elbows, nose etc.) using different heuristics and algorithms including texture matching, edge detection, profile matching and other.

Analysis of body posture, gestures, hand and body movements is also an active research area in many computer vision applications. For example, in [16], three types of affective states are recognized using combined sensory information from the face video, the posture sensor and the game being played. In [17] facial expressions are combined with hand gestures for recognition of six prototypical emotions.

2.2 Emotion Recognition from Audio Channel

Apart from facial expression, human emotions can also be indicated by voice intonation and also by spoken words. Voice intonation expressed in its pitch, tone

and cadence may be described by a set of parameters calculated for time and spectral domains in subsequent time windows. Some sample parameters include signal amplitude, mean and power value, zero-crossing as well as Melfrequency Cepstral Coefficients (MFCC) which has been widely used in acoustic signal processing as they fully consider the hearing characteristics of human beings [19]. In most case the audio channel is used in multimodal affect recognition systems combined with facial expression analysis [22,8].

2.3 Emotion Recognition from Physiological Signals

Cognitive-motivational-emotive model assumes that human affect expression consists both of emotional state contains both of affect as well as physiological response. Sometimes affect expressions may be better or worse controlled by human but it is very hard or even impossible to control physiological reactions.

There are many physiological signals that can be measured and input to the system as indicator of human emotions. The most important signals are: skin conductance and temperature, blood volume pulse, electromyography, respiratory signal and electrocardiography. All the physiological sensors are non-invasive but sometimes they may be intrusive or not comfortable for users. For evaluating peoples emotions by physiological signals special equipment is needed.

There are two main problems with using of physiological signals in automated emotion recognition. Firstly, most of these signals respond also to other inner and outer factors such as human health and physical condition, temperature, humidity and others.

Secondly, measuring physiological signals often cause some inconvenience to the user what is rather unacceptable in a usual computer usage. Fortunately, a great development of non-intrusive, miniaturized and low power remote sensors has been observed that allows for remotely collecting behavior and physiological data from people. Also several computer hardware manufacturers plan to embed some physiological sensors into the control devices for video games.

2.4 Enhancing Emotion Recognition Using Scene Depth Data

The common problem of algorithms processing visual (both achromatic and color) data are insufficient and uneven lighting conditions. Though, there are many attempts to correct scene luminance in such case, they are not always sufficient enough. Depth sensors allow for acquiring the depth image of a scene. As the most popular depth sensors use non-visual infrared light technology they are generally resistant to common problems of RGB cameras. That is why information from depth sensors seems to be very useful when combined with optical channel or even not.

Depth sensors have become very popular since introducing Microsoft Kinect sensor for Xbox console in 2010. This relatively cheap sensor, accompanied with RGB camera, set of microphones and natural user interface (NUI) library gives to software developers a powerful tool for creation of applications that understandİ

human stature, movement, and speech, and in the nearest future will try to recognize user face and emotions as well.

Depth sensors can be used to enhance processing of visual information at different stages of recognition pipeline (Fig. 1). For example, initial test indicate that scene depth information significantly improve image segmentation allowing to extract users hand and body from the background [13]. Also efficiency improvement for face localization and recognition of characteristic face points is expected due to the 3D nature of human face. The theoretically estimated and practically measured Kinect resolution is about 2mm per voxel and seems enough for mentioned applications. There is also a specific problem with processing of depth information which is noise level. Initial tests show that it can be efficiently limited using low pass filtering in time domain with a window of several frames [13].

3 Preliminary Results and Future Plans

The main goal of the research project Emotion Recognition for Affect-Aware Video Games (ERAAVG) is creation of a complete multi-modal affect recognition system to provide automatic and continuous affect recognition of video game players and definition of a complete framework that would enable cooperation between developed affect recognition system and any video game by application programming interface (API) defined within the project. To validate results of the project a prototype affect aware video game will be developed that would adopt its parameters, such as difficulty level, pace or humor basing on the current affect state of a player and challenges in the game.

The first task of the project is to create a multimodal database of emotions recordings that will serve as learning and testing sets for classifiers developed in succeeding steps. The second task of the project is to work out proper preprocessing and segmentation techniques as well as to define the most informative features sets for each information channel. The third task of the project is to create new algorithm for facial expression recognition basing on RGB and depth images. The fourth task of the project is to create separate algorithms for affect recognition basing on data from different channels. In order to validate results of emotion recognition we plan to develop a prototype game with bidirectional interface to affect recognition software.

In the following subsections the preliminary results of ERAAVG project are described.

3.1 FEEDB Database

Facial Expressions and Emotions Database (FEEDB) has been created as the first task of the project [13], [14]. This multimodal database contains totally 3200 recordings of emotions that will serve as learning and testing sets for classifiers. There are two parts, or versions, of FEEDB. The first one consists of 1650 recordings of 50 persons posing for 33 different facial expressions and emotions [13] which are stored using Microsoft proprietary XED format. The second

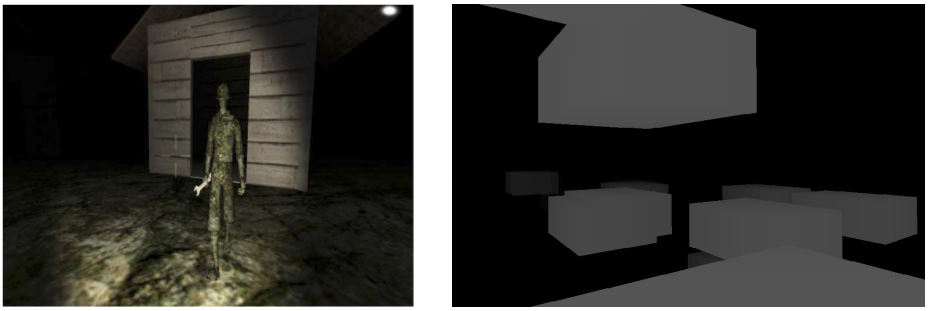
version of FEEDB consists of 1550 recordings of 50 persons [14] recorded as AVI video streams. Additionally, the most interesting set of 9 emotions had been selected to be expressed spontaneously. Special video materials had been prepared to provoke tested persons to naturally express the following expressions: surprise, joy, anger, scorn, disgust, sadness, fear, concentration, and excitement.

The database consist of synchronized multimodal recordings with RGB, depth, and audio channels acquired using Microsoft Kinect for PC sensor at a resolution 640x480 pixels at 30 fps. Some of the recordings are accompanied by additional information from physiological signals (heart rate, skin conductance, breath and temperature) acquired using FlexComp Infinity device with appropriate sensor set. Chosen frames of recordings are indexed by defining characteristic points of the face and can be used for learning and testing of the classifier for emotion recognition.

3.2 Face Detection and Emotion Recognition

FEEDB database has been used as a learning and testing dataset in a prototype system for recognition of 9 emotions using recordings of depth channel only. The set of emotions covers neutral, joy, surprise neutral and positive, euphoria, fear, fright, anger, and scorn (Fig. 2). The application locates the face and its characteristic points, then classifies emotions on the base of recognized action units (fundamental movements of face muscles in Facial Action Coding System). The system recognizes emotions in real-time and its average efficiency of the recognition for 25 persons was 50%. Though it is not very high it is a good starting point for further research especially in fusion of ther input channels. The experiment proved that FEEDB can be practically used in face and emotion recognition tasks.

Fig. 2. Application for emotion recognition using depth maps



(a) (b)

Fig. 3. Sample screens from prototype of an affect-aware video game (a) Zombie land (b) moving platforms

3.3 Affect Aware Video Game

A prototype affect aware video game has also been developed within ERAAVG project using Unity 3D game engine (Fig. 3). The game consists of three levels with different challenges, namely escaping from attacking zombies (Fig. 3a), jumping across moving platforms (Fig. 3b) and finding exit of dynamically reconfiguring maze. The difficulty of each level changes depending on current player state. For example at the first level the user under the stress finds torch batteries more frequently to threaten zombies. In turn, the speed of movings platforms depends on user self-assurance, but this time platforms speeds-up to make the game more difficult. Unfortunately, the game has not been connected with emotion recognition application and predicts users states basing on users behavior only. Nevertheless, initial experiments with several players has resulted with positive opinions as the game was able to surprise the player by dynamic changes.

3.4 Conclusions and Future Works

Affect aware video games seem to be very interesting to players as they engage players emotionally and can dynamically react to players emotional state which was confirmed by a prototype game developed within ERAAVG project. Such video games can have an entertainment character but can also be used in other applications, such as e-learning, psychological training or therapy, and even in marketing systems.

Automatic emotion recognition is a key to create affect-aware software. A prototype emotion recognition system proved that real-time emotion recognition is possible, though fusion of many input channels is needed to receive reliable results. A comprehensive data set, such as FEEDB, is also needed to develop emotion recognition system.

The next task in ERAAVG project is to develop a bidirectional framework that could connect affect recognition application with any affect-aware video

game using a specific application programming interface. Further steps cover development of more efficient affect recognition algorithm using multimodal data such as RGB and depth video, audio channel, and physiological sensors. Also other affect aware video games will be developed that could increase players satisfaction from a gameplay.

Acknowledgment. The research leading to these results has received funding from the Polish-Norwegian Research Programme operated by the National Centre for Research and Development under the Norwegian Financial Mechanism 2009-2014 in the frame of Project Contract No Pol-Nor/210629/51/2013.

References

1. Derek, J., Sollenberger, Â., Munindar, P.: Singh: Koko: an architecture for affect-aware games. In: Proceedings of the First International Workshop on Agents for Games and Simulations. Springer (2010)
2. Liu, X., Zhang, L., Yadegar, J.: An intelligent multi-modal affect recognition system for persistent and non-invasive personal health monitoring. In: 2011 IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pp. 2163–2167. IEEE (2011)
3. Picard, R.: Affective Computing: From Laughter to IEEE. IEEE Transactions on Affective Computing 1(1) (2010)
4. Adams, E.: Fundamentals of Game Design. New Riders Publishing (2009)
5. Obaid, M., Han, C., Billinghurst, M.: Feed The Fish: An Affect-Aware Game. In: Australasian Conference on Interactive Entertainment 2008, Brisbane, Australia, December 3-4 (2008)
6. Leadbetter, R.: Games will detect your feelings (2011), `http://www.eurogamer.net/articles/digitalfoundryvalve-biometrics-blog-entry`
7. Russell, J.A., Mehrabian, A.: Evidence for a three-factor theory of emotions. Journal of Research in Personality, 273–294 (1977)
8. Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Cambridge University Press, Cambridge (1988)
9. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 22(12), 1424–1445 (2000)
10. Fasel, B., Luettin, J.: Automatic Facial Expression Analysis: A Survey. Pattern Recognition 36(1), 259–275 (2003)
11. Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan, S.: Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information. In: Proc. of ACM 6th International Conference on Mutlmodal Interfaces (2004)
12. Gunes, H., Piccardi, M.: Affect Recognition from Face and Body: Early Fusion vs. Late Fusion. In: Proc. IEEE Int. Conf. Systems, Man, and Cybernetics SMC, pp. 3437–3443 (2005)
13. Szwoch, M.: On Facial Expressions and Emotions RGB-D Database, BDAS (2014)
14. Szwoch, M.: FEEDB: a multimodal database of facial expressions and emotions. In: Proc. of the 6th Int. Conf. on Human System Interaction, pp. 524–531 (2013)

15. Pantic, M., Rothkrantz, L.J.M.: Toward an Affect-Sensitive Multimodal Human-Computer Interaction. Proc. of IEEE 91(9), 1370–1390 (2003)
16. Kapoor, A., Picard, R.W., Ivanov, Y.: Probabilistic Combination of Multiple Modalities to Detect Interest. In: Proc. IEEE ICPR (2004)
17. Balomenos, T., Raouzaiou, A., Ioannou, S., Drosopoulos, A., Karpouzis, K., Kollias, S.D.: Emotion Analysis in Man-Machine Interaction Systems. In: Bengio, S., Bourlard, H. (eds.) MLMI 2004. LNCS, vol. 3361, pp. 318–328. Springer, Heidelberg (2005)
18. Ekman, P., Friesen, W.V.: Unmasking the face: a guide to recognizing emotions from facial clues. Prentice-Hall, Imprint Englewood Cliffs (1975)
19. Sigurdsson, S., Petersen, K., Schioler, T.: Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music. In: Proc. Int. Conf. Music Inf. Retrieval, pp. 286–289 (2006)
20. Chen, C.W., Wang, C.C.: 3D Active Appearance Model for Aligning Faces in 2D Images. In: International Conference on Intelligent Robots and Systems IROS, pp. 3133–3139 (2008)
21. Xiao, J., Baker, S., Matthews, I., Kanade, T.: Real-Time Combined 2D+3D Active Appearance Models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 535–542 (2004)
22. Chen, L.S., et al.: Emotion recognition from audiovisual information. In: IEEE Second Workshop on Multimedia Signal Processing, December 7-9, pp. 83–88 (1998)

Registration and Analysis of Data during Registration and Exposure Process

Jarosław Zdrojewski[1], Adam Marchewka[1], and Rocío Pérez de Prado[2]

[1] Institute of Telecommunication and Informatics,
University of Technology and Life Sciences, Prof. S. Kaliskiego 7, 85-791 Bydgoszcz
{jaz,adimar}@utp.edu.pl
[2] Telecommunication Engineering Dpt., University of Jaén, Alfonso X El Sabio 28, Linares, Spain
rperez@ujaen.es

Abstract. The article presents printing system used for printed circuit board production. The system main task is an image transfer from the film to the resist covering the panel. Beside the complexity of film and panel alignment based on predefined registration marks, cameras and sophisticated algorithms, system can be considered as a solid source of production feedback. Proposition of strategy, parameters selection, database organization, and analyze tools possibilities are presented together with experimental results showing capability and limits of the proposed system.

1 Introduction

The printed circuit board manufacturing process has a task of creating complex electronic circuits using simple ideas and technique. This fairly simple process of PCB production includes many steps and some of them may became complex when high precision is one of the requirements. The board is a base for electronic components assembly. It realizes electrical connections. Each connection should be solid and should fallow specified parameters. When reviewing from the side of rigid result then crucial are position, width, thickness of tracks and stability of innerlayer connections using plated holes. Manufacturing cycle begins with the project files including specification sent to the producer. It is conducted in a technological process consisting many stages (Fig. 1), where printing is one of the first ones. The whole process gets even more complicated when some connections in the multilayer printed circuit board are made using buried or blind holes. The majority of these stages require continuous control, which is in most cases conducted automatically in optical measuring devices.

Geometry and electrical parameters of constructed circuit is measured to keep observed deviations within defined tolerance limits in order to reach the quality of the final product through applying required corrections to the technological process. Usually producer dream will be to get solid feedback from each step of process and after applying some efforts to analyze data get to the point where analyze results can lead to the process improvement by increasing quality of the

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_29

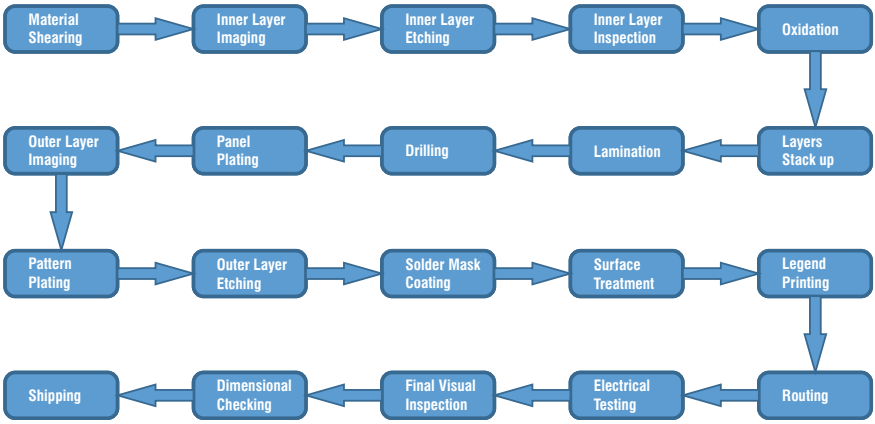


Fig. 1. Simplified Printed circuit board manufacturing process flow

product or production process (reducing costs or optimizing time or minimizing the risk of failure)

Printing has a task to create image on the panel representing a layer of tracks. Modern printers have inbuilt ability to maintain registration (film positioning in relation to panel) and imaging task by using complex sensor like digital cameras to take some measurements and use those values to optimize process and printing quality. Results representing such system are shown on Fig. 2. In this case printer is equipped with 4 digital cameras - image processing results take left part of the screen - the right one is presenting the registration results including films/panel deviations and alignment final result.

2 Exposure Process and Its Complexity

Algorithms for processing data based on images from multi camera device is complex (Fig. 3). All data inputs are based on image processing using dedicated targets placed on the films and panel [2]. Results are applied to inputs of sophisticated mechatronic device responsible for motion.

System realizing alignment and imaging task can be considered on a different abstract levels. In general we can divide it into areas serving basic function of the unit, currently realized process data, inspecting and debugging, error logging, system optimization data. In general we have 5 groups:

- Printer hardware data (axis location, mechanical stability, Images quality),
- Job settings, registration and printing parameters,
- Registration results (films/panel deviations, alignment data, alignment error),
- Error log (including system or process failures, errors or warnings)
- Optimization data (automatically saved pictures used to optimize system stability and response)

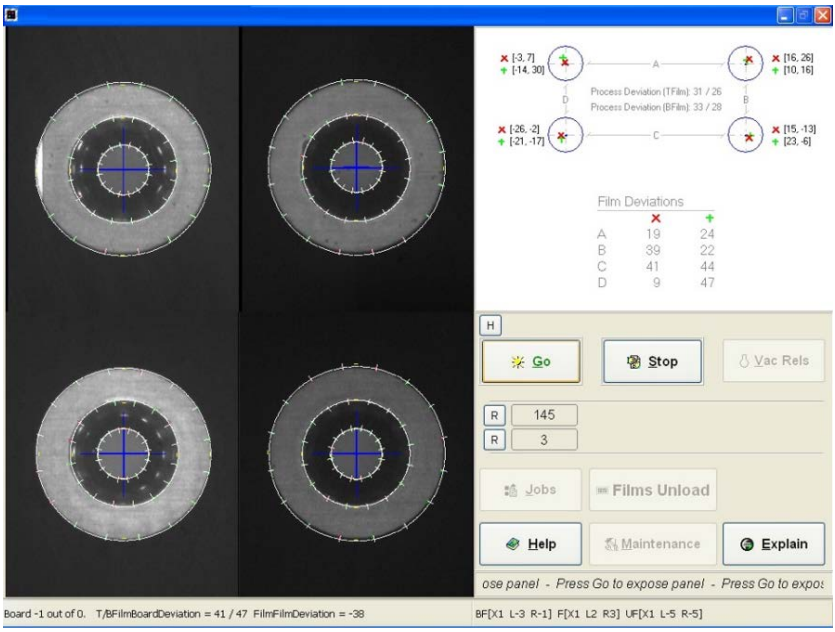


Fig. 2. Registration status - films deviation and alignment results

In the group of printer hardware we have part strictly related to system configuration. For sure we have to count with 2 axis for each camera, and 3 axis for motion part representing alignment for a given layer.

Basically 4 cameras and both films actuator will define 14 axis, where each of them will be connected with coordinate system and will carry X or Y coordinate and some parameters like hysteresis, speed, acceleration or offset. Job setting of course have task to define job conditions for a mentioned above hardware and job specific information, like alignment type, registration targets, process tolerances, light conditions for the cameras, and a way how to use information from the camera for a registration.

Registration is a part of the printing process realized just before exposure. Printing will take a place only in the case of film and panel be aligned and relative position of them being in the predefined tolerances (Fig. 2). Nevertheless how sophisticated can be system of defining tolerances - outcome must lay in the strict limits and all differences can be measured and analyzed. This part has a big impact on the board quality because for a human eye is easy to see if tracks patter matches pattern of drilled in the panel holes if not then how much. General goal of the system during alignment based on the input from the registration cameras is to minimize deviations and improve also this first sight effect.

Every machine needs also a part responsible for carrying system hardware status and diagnostic information. In this case this task is represented by error

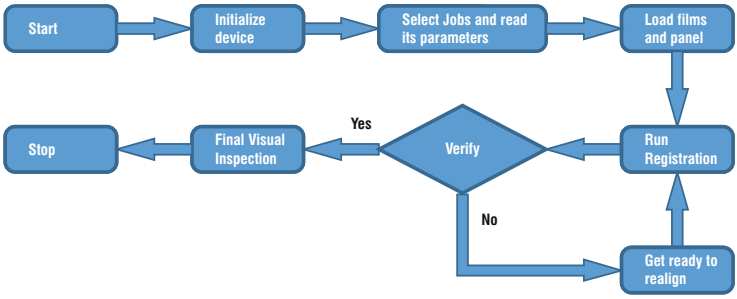


Fig. 3. Registration and imaging - algorithm steps

log holding a data including system or process failures, errors and warnings. This part can be used to verify system stability or making decision what action take to fix a unit in the case of failure.

Significant part is a pert representing optimization data. Very often system has to deal with a new type of material defining new, difficult conditions for a vision system. In this case system can automatically save images and they can be used to optimize offline system stability and response.

In general, considering imaging system complexity problem to solve will be to define, implement and evaluate complexity of system data managing software module.

3 Data Managing Tool

Printing unit required data can be saved in the database. Its structure has to reflect structure of the system, using traditional technics of design for the relational data base. In general data is divided in to 3 parts:

- Registration results (Results table),
- Error log (ErrorLog table),
- Optimization data (Optimization table).

Each part of the system is represented by separate table (Fig. 4), with a structure representing related unit part and function of the unit to model.

```
CREATE TABLE Results (
 ID INTEGER,
 JobName CHAR(64),
 WorkOrderNumber CHAR(64),
 …
 PRIMARY KEY (ID)
)
```

Fig. 4. SQL query to create table defining Results database to keep registration data

Number of values to save has to be reflected in the application used to restore and process information from a database. Sample screens showing the interface structure is shown on the Fig. 5. User can pick a Job name, work order number related tray and time frame before next step defining what part of data is going to be extracted. In relation to the data part optional element include: job name, work order number, operator, counter, film deviations, dimensional tolerance, alignment results, alignment tolerances, tray info and date and time mark to find out when the record was placed in the database.

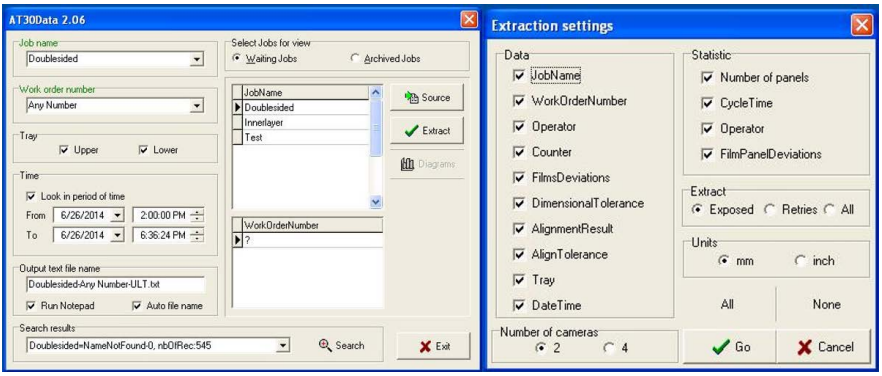


Fig. 5. Settings used to define area of interest when extracting data. Each row can be included and can be related to the desired stage of the process.

Beside data part also some basic statistic can be calculated, related to the total number of panels or related to the operator. In this group film/panel deviation statistic is included as well. Extracted data can represent final alignment approach or intermediate steps preceding final result.

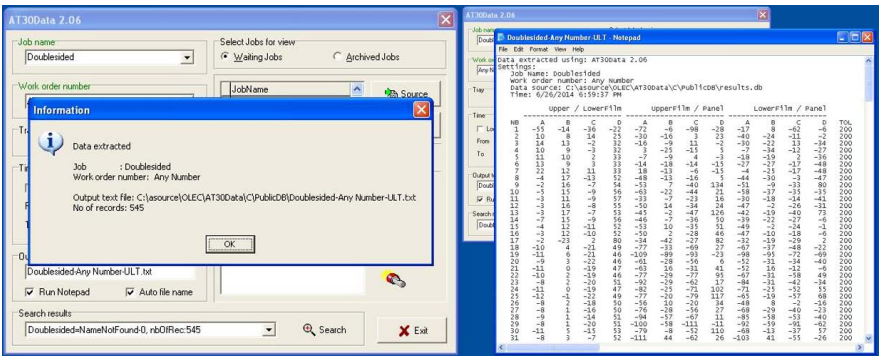


Fig. 6. Result of data extracting- amount of data strictly depends on extracting tool settings including time frame defining limits for analyzed records

Data extraction will basically produce text file (Fig. 6) including all desired information preselected in the initial step. In this case resulting text file has to be analyzed by additional tools, like EXCEL simply by importing functions. In this case all needed calculation has to be defined using formulas specific for the program. All columns and rows of data are clearly to identify by dedicated label or row number in this case representing number of panel in group belonging to the given analyzed Job.

4 Experimental Results

Main purpose of the experiment is verification of system ability to supply desired structure of a process data and its capability to be represented in appropriate graphical form.

Database can be a space for huge amount of data. Extracting tool can access predefined portion but the outcome can be not easy to process (Fig. 7). Text format and local view can create some difficulties to make easy and proper decision. Text results file include: deviations for upper / lower film, upper film / panel, lower film / panel, alignment results represented by distances bottom film to Panel, upper film to panel, upper film to bottom film, tray location, date, time, job name, work order number, operator. Also some statistical data is included.

```
Data extracted using: AT30Data 2.06
Settings:
   Job Name: 1500
   Work order number: ?
   Data source: C:\ASource\OLEC\AT30Data\C\GCE\1\PublicDB\results.db
   Time: 2013-10-22 05:54:54
   From: 2013-10-21 14:00:00
   To  : 2013-10-23 05:54:48

          Upper / LowerFilm       UpperFilm / Panel       LowerFilm / Panel
     ------------------------ ------------------------ ------------------------
   NB     A     B     C     D     A     B     C     D     A     B     C     D
 1163    95    -3     5   -49    95    -3   109   -20     0     0   104    29
 1163    89     6     6   -50    89     6   108   -23     0     0   102    27

                                Bottom2Panel                                                
```

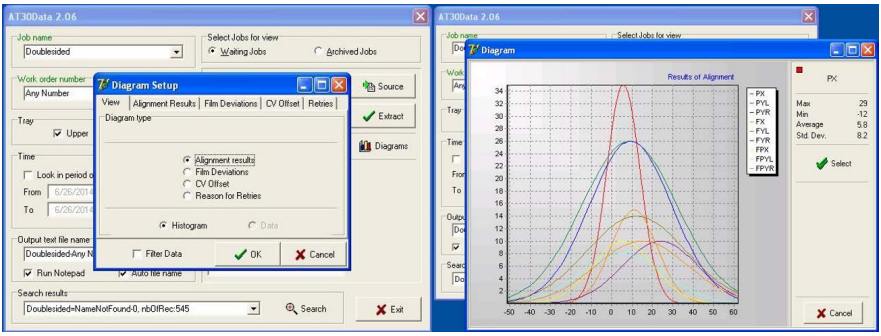


Fig. 8. Processing exposure process data - graphical presentation for statistics of registration results. Each curve is representing separate axis behavior and its base limits.

5 Conclusions

Printing process can generate huge amount of data. To find it useful, right tools are necessary and we need to create flexible strategies to extract information. Efficient analyze will be possible only if data extraction will be followed by statistical analyze. In this case software tool has to supply graphical interface to simplify understanding and speed up viewing. In most cases system will use local database and remote access can be necessary not only to supervise process but also to maintain and serve system stability and optimization. The paper describes exposure system data maintaining using dedicated data structure. The main objective of this article is to present complexity of task and its needs to be fully customized for a given system. For the future work presented system will be supplemented by its extensions for serving remote access and data transfer for automatic optimization and early failure and instability detection.

References

1. Coombs, C.F.J.: Printed Circuit Handbook, 6th edn. McGraw-Hill (2007)
2. Marchewka, A., Zdrojewski, J.: Film Alignment Before Solder Mask Exposure. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 3. AISC, vol. 102, pp. 387–393. Springer, Heidelberg (2011)
3. Ohlig, B.: Registration and Knowledge Through and Across Your Panel, CircuiTree (April 1, 2002)
4. Wocianiec, R., Zdrojewski, J.: Film Position Optimization Before Exposure. Image Processing & Communication An International Journal 5(2), 35–46 (2010)

Detection of Defects in Carbon-Fiber Composites Using Computer-Vision-Based Processing of Microwave Maps

Bartłomiej Zieliński[1], Marcin Iwanowski[1], Bartłomiej Salski[2], and Szymon Reszewicz[1,2]

[1] Institute of Control and Industrial Electronics
Warsaw University of Technology
ul. Koszykowa 75; 00-662 Warszawa, Poland
{zielinsb,iwanowski}@ee.pw.edu.pl
[2] Institute of Radioelectronics, Warsaw University of Technology
ul. Nowowiejska 15/19; 00-665 Warszawa, Poland
bsalski@ire.pw.edu.pl

Abstract. In this paper, a method for analyzing data gathered by a non-destructive inspecting approach of carbon-fiber composites is presented. In the process of composite probing a map of measurements is obtained. Upon the measurements an analysis can be performed in order to decide if the composite under examination is defective or not. The map is composed of a set of separate probes, which may be viewed as a set of pixels and, therefore, whole map can be perceived as an image. For this reason, image-based processing has been applied to examine the maps. The proposed method allows to detect suspicious regions with different, user-defined sensitivity.

1 Introduction

The non-destructive defectoscopy [1,3,4,7,9] plays an important role in the monitoring of constructions. Among the construction materials recently used to build ships, planes and other vehicles, the composites became one of the most popular. The main reason for such popularity is their low weight combined with high strength and ease of forming. The composite materials, however, may suffer from different types of damages, which can appear both in the production process and during their usage. When external layers of the composite are destroyed, it is easy to discover the problem. However, some damages can unfortunately be located in the internal layers of the composite. Then, inspection is much more difficult and uncertain as well. Dubious areas can obviously be controlled in a destructive way. On the other hand, it is possible to perform hidden composite layer inspecting in search for defective regions. In this paper, we present a method of analyzing maps acquired by non-destructive composite probing in order to spot defect suspicious areas.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_30

2 Radio-Frequency Scanning

The measurements are performed by means of radio-frequency sensing with use of specially designed sensor connected to a micro-controller. Materials considered in this research are multi-layer carbon-fiber-reinforced polymer composites. They, as such, are highly conductive. It is, therefore, possible to examine the material with use of magnetic field which penetrates the composite. For this purpose a special sensor has been constructed and evaluated. It is composed of two coupled spiral inductor coils. One is driven by a radio-frequency signal, and the other is connected to the power detector. The sensor induces eddy currents [2,6,9] inside the composite. This way the sensor makes possible to measure transmission coefficient S_{21}, which is dependent on conductivity in the examined composite. The conductivity, in turn, changes when there are defects in the composite—the presence of defects in the inspected material influences the amount of the eddy currents. The example of the coefficient as a function of frequency is presented in the Fig. 1.

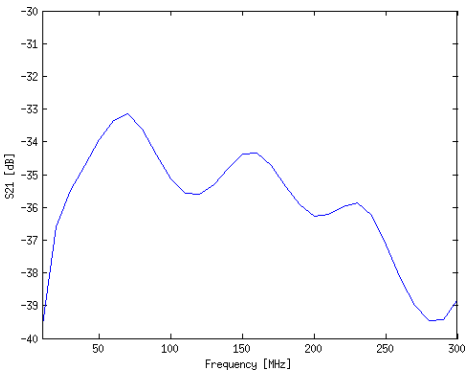


Fig. 1. Exemplary spectrum of a transmission coefficient S_{21}

3 Proposed Processing Method

A single measurement shot at a given position consists of a transmission coefficient vector for consecutive frequencies (defined in advance by the user). In other words, every position of the sensor yields a vector. Having performed numerous measurements for different sensor positions on examined composite samples, one achieves a three-dimensional matrix. Each position corresponds with a frequency vector. Consequently, the results might be considered as a cube, with two spatial dimensions and the third dimension being probing frequencies. For each frequency one can perceive the transmission coefficients measured at certain positions as pixels, which constitute an image. Therefore, it is possible to analyze the data with use of image processing techniques.

The radio-frequency matrix is, thus, processed in order to find the component defects. The data is composed of 2D matrices of measurement values. Each matrix contains values measured for a different frequency f. Due to this fact, the data can be perceived as a set of 2D images I_f. An example of such set of images—presenting component sample of size around 20×20 cm^2—is shown in Fig. 2. The images are of resolution 1 mm in both axes and present 6 bands for various frequencies.

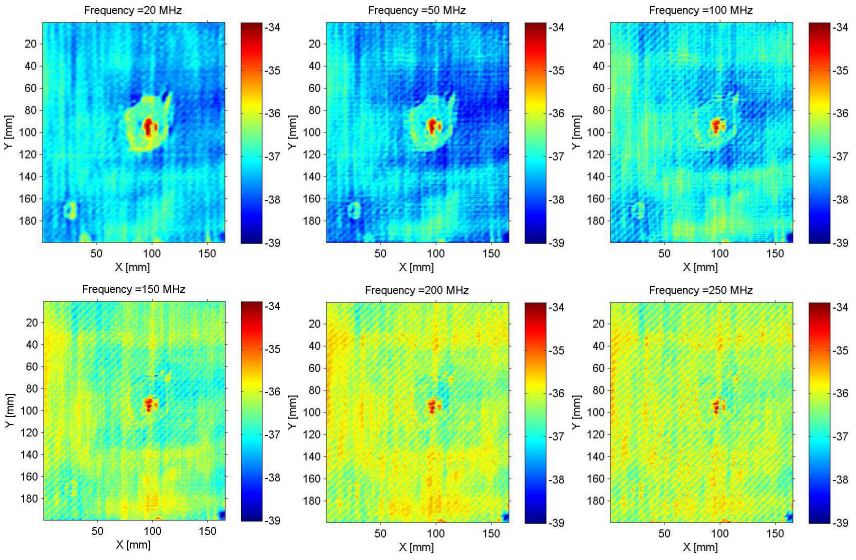


Fig. 2. Example band images at consecutive frequencies (pseudo-colored)

The processing is performed in two-step approach. At first the individual processing of single frequency bands is performed. In the second step the results of the former one are combined in order to produce the final defect map.

3.1 Frequency Bands Processing

The band images I_f might be contaminated with noise, which in this case refers to single measurements with values remarkably different from their neighbors. In order to remove the noise the convolution filter is applied. Defects of the examined composite can appear as groups of connected pixels of values considerably different than average pixel value in an image. To mark these dubious pixels binarization by thresholding is applied. Every band image is treated separately—for every image the threshold values are computed based on the image content. In effect, the thresholds may differ from one image to another. In order to find the optimal threshold values, for each image I_f, the median value of all pixels is calculated. It constitutes the most probable value robust against

presence of extreme ones [5], which might be delivered by e.g. measurement failures. The suspicious pixels have significantly higher or lower values than the pixels of undamaged regions. Thus, the next step is comparing the pixel values with predefined thresholds:

$$\forall_p J_f(p) = \begin{cases} 1 \text{ when } & (m - \Delta_f^-) < I(p) < (m + \Delta_f^+) \\ 0 \text{ otherwise,} & \end{cases}, \tag{1}$$

where p stands for pixel coordinates, $J_f(p)$ is an outcome image, $m = med\{I_f\}$ stands for median value of pixels of I_f, and (Δ_f^-, Δ_f^+) are lower and upper margins for frequency f. As a result, we produce mask-images, where 0-valued pixels correspond to pixels of values greater or smaller than the median by the threshold values in the input images. This way, for each frequency, an image of possible defects is received.

For various frequencies f the thresholds (Δ_f^-, Δ_f^+), expressed in [dB], might differ. This is because of possible physical properties of material under test. For different frequencies there can be a necessity of applying unlike thresholds. Moreover some frequencies might be more relevant then the others.

3.2 Defect Map Generation

As a result of the previous processing step the set of binary images representing the possible defects at various frequencies is obtained. It might happen that the defects are visible only on some frequencies, not on all images I_f. In order to gather the information from all frequencies and to get single defect image, we calculate a weighted sum of binary images:

$$S = \sum_{\forall f} w_f \cdot J_f, \tag{2}$$

where w_f stands for the weight associated with the frequency f. In effect, a scalar image S of suspicious areas arise. The greater value of particular pixel is, the greater is the chance that the pixel depicts a defected area. This image is a multi-valued defect map, that is used next to determine the strict boundaries of defects. Thus, the final step is thresholding this image, which allows classifying measurements into two classes: defect and non-defect. Depending on the threshold value one gets higher or lower sensitivity of detection. In case only large area defects are expected to find, the classification result may be further processed in order to filter-out small defects with use of e.g. morphological opening filter or opening by reconstruction [8]:

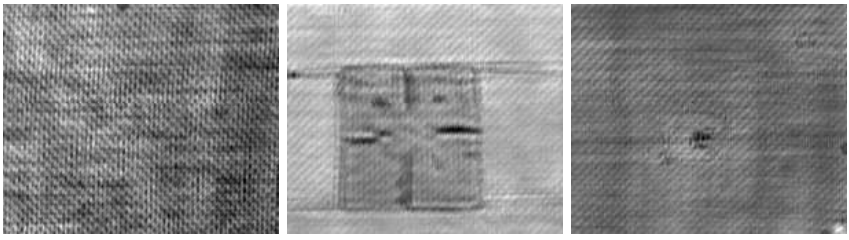
$$D = (S > t) \circ_{N8} B, \tag{3}$$

where D stands for the final defect image and $\circ_{N8} B$ signifies the opening by reconstruction when the structuring element of erosion equals B and the one of reconstruction equals $N8$ (8-connected closest neighbors, simplest 3×3 neighborhood). The choice of B depends on the size of unwanted objects that should

be removed from the final defect map—it should be smaller than the smallest expected defect. The parameter t stands for the threshold applied on image S. In case of all weights $w_f = 1$ (equal weight for each band), the value of t represents the number of bands (frequencies) on which the defect was detected.

4 Results

The presented method has been evaluated on measurements taken for carbon-fiber-reinforced polymer component samples. One of them was undamaged sample, the other two were defective—the first one had a "crack" type damage and the second had a "hole" type damage. Probing frequencies equaled $f \in \{20, 50, 100, 150, 200, 250\}$ MHz. The measurement matrices for frequency $f = 200$ MHz are shown in the Fig. 3.

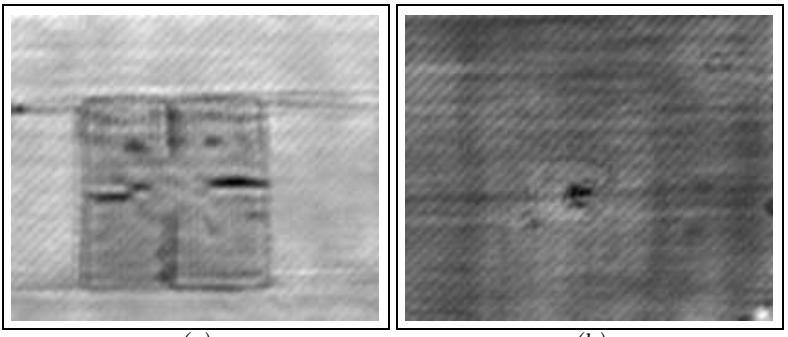


(a) *(b)* *(c)*

Fig. 3. Radio maps of: (a) undamaged sample, (b) "crack" type damage, and (c) "hole" type damage, for frequency $f = 200$ MHz

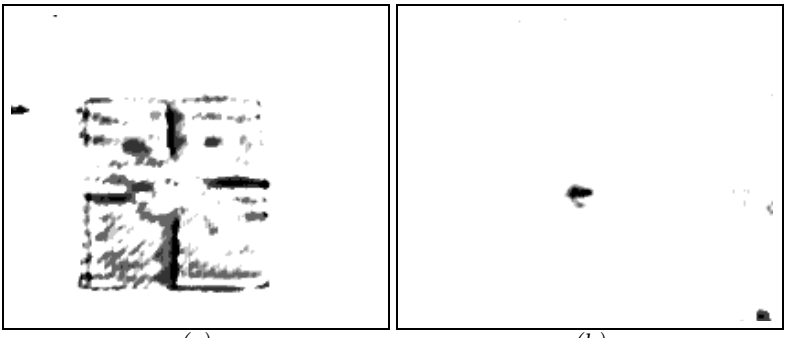
The first step of processing was filtering. Every matrix, perceived as an image, was filtered with local averaging filter in order to remove noisy measurements (Fig. 4). In our case the best results were obtained with use of the local averaging filter with mask of size 5x5. Then, images were thresholded. All values greater or smaller than the median by a certain thresholds were marked as dubious. In this way, for every composite sample, a set of images arose. Next, a weighted sum of the images was calculated (Fig. 5). The final phase was thresholding the obtained sum. In effect, an image of suspected defects was achieved (Fig. 6). It is worth mentioning, that the image might be subjected to filtering, i.e. morphological opening, in order to remove defects of insignificant size, so that only damage regions of certain area remain (Fig. 7).

In all the experiments a pixel was marked as suspicious for a certain frequency map, if its value was greater than 3 dB or smaller than 2 dB than the median. All the frequency components were treated with the same relevance. The final threshold equaled 50%, which means that a pixel was assessed as defective, if it was marked as suspicious in at least 50% of frequency maps. Successive steps of image processing are presented in Fig. 3—7 (in undamaged composite sample



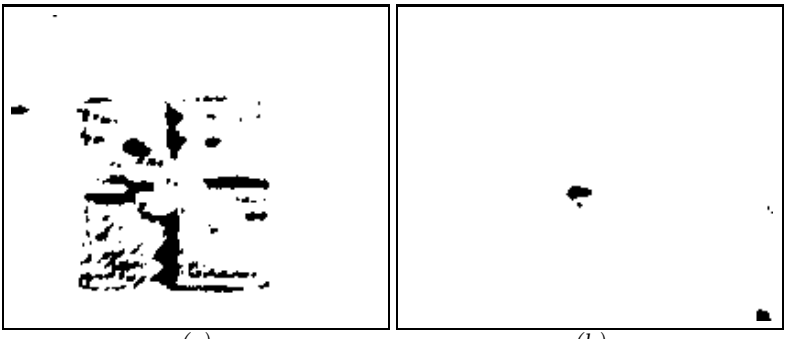
(a) (b)

Fig. 4. Radio maps processed with averaging filter of: (a) "crack" type damage, and (b) "hole" type damage, for frequency $f = 200$ MHz



(a) (b)

Fig. 5. Sum of thresholded images of: (a) "crack" type damage, and (b) "hole" type damage



(a) (b)

Fig. 6. Defective regions of: (a) "crack" type damage, and (b) "hole" type damage

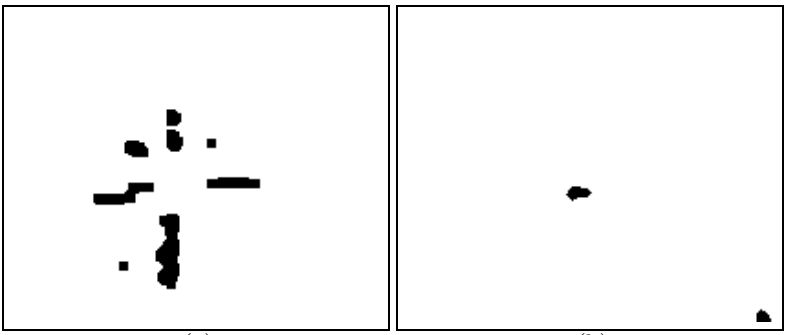


Fig. 7. Morphological opening of defective regions of: (a) "crack" type damage, and (b) "hole" type damage

no defects were detected, so only damaged samples are interesting and presented in the figures). As one can see, the automatic defect classification gives correct results—the defected regions are clearly marked—for both "crack" and "hole" type damage samples.

5 Conclusions

Non-destructive composite probing is an important trend in material examination and gain more and more interest in the field of controlling material condition. Unlike as with use of invasive methods, composites can be scanned, so that a vector of values is obtained for each measured position. These positions may be viewed as pixels and, therefore, set of separate measurements can be perceived as an image. In this paper, a method of analyzing such images was presented. This method consists of two steps—at first the frequency band are processed individually, next the final classification is performed. User can set various thresholds and, in effect, achieve an image, where defective regions are marked. The method was tested on a series of image sets and achieved high efficiency and accuracy.

Acknowledgments. The research leading to these results has received funding from EU 7th Framework Programme managed by REA under grant agreement no. 314935. Part of this work was also funded by the Polish National Ministry of Science and Higher Education under grant agreement no. 2826/7.PR/2013/2. The research has been undertaken as a part of the project entitled Radio Frequency Sensing for Non-Destructive Testing of Carbon Fibre Reinforced Composite Materials for Structural Health Monitoring Comp-Health, in the collaboration between the following organizations: E.T.S. Sistemi Industriali srl, Kingston Computer Consultancy Limited, Nemetschek ood, Atard a.s., UAB Elmika, Warsaw University of Technology, National Technical University of Athens and Brunel Innovation Centre of Brunel University.

References

1. Bui, H., Wasselynck, G., Trichet, D., Fouladgar, J., Ramdane, B., Berthiau, G.: Thermo inductive nondestructive testing method applied to cfrp. COMPEL -The International Journal for Computation and Mathematics in Electrical and Electronic Engineering 33(1-2), 167–180 (2014)
2. Cheng, L., Tian, G.: Surface crack detection for carbon fiber reinforced plastic (cfrp) materials using pulsed eddy current thermography. IEEE Sensors Journal 11(12), 3261–3268 (2011)
3. Cheng, L., Tian, G.: Comparison of nondestructive testing methods on detection of delaminations in composites. Journal of Sensors (2012)
4. Hung, Y., Chen, Y., Ng, S., Liu, L., Huang, Y., Luk, B., Ip, R., Wu, C., Chung, P.: Review and comparison of shearography and active thermography for nondestructive evaluation. Materials Science and Engineering R: Reports 64(5-6), 73–112 (2009)
5. Meer, P.: Robust techniques for computer vision. In: Emerging Topics in Computer Vision, pp. 107–190 (2004)
6. Nabavi, M., Nihtianov, S.: Design strategies for eddy-current displacement sensor systems: Review and recommendations. IEEE Sensors Journal 12(12), 3346–3355 (2012)
7. Scott, I., Scala, C.: A review of non-destructive testing of composite materials. NDT International 15(2), 75–86 (1982)
8. Soille, P.: Morphological Image Analysis. Springer (1999)
9. Sophian, A., Tian, G., Taylor, D., Rudlin, J.: Electromagnetic and eddy currentndt: A review. Insight: Non-Destructive Testing and Condition Monitoring 43(5), 302–306 (2001)

Part II

Communications

Network Anomaly Detection Based on ARFIMA Model

Tomasz Andrysiak and Łukasz Saganowski

Institute of Telecommunications, University of Technology & Life Sciences in Bydgoszcz ul. Kaliskiego 7, 85-789 Bydgoszcz, Poland
{tomasz.andrysiak,lukasz.saganowski}@utp.edu.pl

Abstract. In this paper, the estimation model ARFIMA is presented as a method of detecting anomalies in network traffic. Parameters estimation and model identification are performed with the use of algorithms of: Geweke and Porter-Hudak (estimation of the differencing parameters) and Box-Jankins (identification of the row of the model). The choice of optimal parameters of the model is realized by a compromise between model consistency and size of its estimation error. In the proposed method, we use statistical relations between estimated traffic model and its real variation to detect its abnormal behaviour. The obtained experiment results confirm effectiveness of the presented method.

1 Introduction

Intrusion Detection/ Prevention Systems (IDS/ IPS) are currently one of the main mechanisms of overseeing safety of computer networks. Their aim is to identify, detect and respond to an unauthorised activity directed towards protected network resources. IDS Systems can be divided into two categories depending on different techniques of threat identification. One category is based on detection of known attacks with use of signature techniques. The other utilizes the idea of monitoring normal operation of the system in order to detect its abnormal behaviour, which may be indicative of an intruder. Such an approach allows to reveal intrusion attempts comprising of numerous network connections. Actions based on probing networks and scanning ports are examples of that kind of attacks [1]. The main advantage of the method based on anomaly detection is the ability to recognize unknown attacks disrupting normal network traffic. Therefore, IDS/ IPS systems using anomalies are more effective than signature-based systems in relation to novel and unknown types of attacks. Statistical techniques have found application in Intrusion Detection Systems because of their ability to detect novel intrusions and attacks, which cannot be achieved by signature-based approaches. It has been shown that network traffic presents several relevant statistical properties when analysed at different levels (e.g. self-similarity, long range dependence, entropy variations, etc.). The most commonly applied approach is grounded on methods using detection of statistical anomalies on the basis of estimated specific profiles of network traffic parameters. The profiles often represent the average size of network traffic parameters, such Saganowski:

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_31

the number of IP packets, average number of newly dialed calls within a time unit, the ratio of packets of particular network protocols, etc. Moreover, it can be observed that there is some statistical dependence resulting from the part of a day or week, as well as there can be held statistics for particular network protocols. IDS systems based on the mentioned methods can learn the typical network profile - the process lasts from few days to few weeks - and compare the current activity of the network to the remembered profile. Comparison of the two profiles will constitute foundation for stating whether there is something extraordinary occurring in the network or not [2]. Currently, the most developed methods of anomaly detection are those based on statistical models describing the analysed network traffic as time series. The most common are autoagressive models ARMA or ARIMA, which allow to estimate the characteristics of the analysed network traffic. In the field literature, there are also hybrid methods connecting elements of pre-processing (wavelet decomposition), and afterwards estimating statistical parameters of the processed signal [3]. In this article we present the use of estimation of statistical method ARFIMA for specific time series describing the analysed network traffic. Anomaly detection is realized by estimated parameters of the model, and comparative analysis to network traffic profiles. This paper is organized as follows. After the introduction, in Section 2 the ARFIMA model for date traffic prediction is described in details. Then, in Section 3 the Anomaly Detection System based on ARFIMA model estimation is shown. Conclusion are given thereafter.

2 The Arfima Model for Network Anomaly Detection

2.1 Introduction to ARFIMA Model

The Autoregressive Fractional Integrated Moving Average model called $ARFIMA\ (p, d, q)$ is a combination Fractional Differenced Noise and Auto Regressive Moving Average which is proposed by Grange, Joyeux and Hosking, in order to analysis the Long-Memory property [4].

The $ARFIMA(p, d, q)$ model for time series Y_t is written Saganowski:

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t, \quad t = 1, 2, ...T, \tag{1}$$

where y_t is the time series, $\varepsilon_t \sim (0, \sigma^2)$ is the white noise process with zero mean and variances σ^2, $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$ is the autoregressive polynomial and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_p L^q$ is the moving average polynomial, L is the backward shift operator and $(1 - L)^d$ is the fractional differencing operator given by the following binomial expansion:

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k \tag{2}$$

and

$$\binom{d}{k} (-1)^k = \frac{\Gamma(d+1)(-1)^k}{\Gamma(d-k+1)\Gamma(k+1)} = \frac{\Gamma(-d+k)}{\Gamma(-d)\Gamma(k+1)}. \tag{3}$$

where $\Gamma(*)$ denotes the gamma function and d is the number of differences required to give a stationary series and $(1-L)^d$ is the d^{th} power of the differencing operator. When $d \in (-0,5\ ,\ 0,5)$, the $ARFIMA(p,d,q)$ process is stationary, and if $d \in (0\ ,\ 0,5)$ the process presents long-memory behavior.

In this paper we assume that $\{y_t\}$ is a linear process without a deterministic term. We now define $U_t = (1-L)Y_t$, so that $\{Y_t\}$ is an $ARMA(p,q)$ process. The process defined in Equation 1 is stationary and invertible (see [4]) and its spectral density function, $f_y(\omega)$, is given by

$$f_y(\omega) = f_U(\omega)\left(2\sin\left(\frac{\omega}{2}\right)\right)^{-2d}, \omega \in [-\pi,\pi], \tag{4}$$

where $f_U(\omega)$ is the spectral density function of the process $\{Y_t\}$.

2.2 GPH Estimation of Fractional Differencing Parameter

The GPH estimation procedure is a two-step procedure, which begins with the estimation of d and is based on the following regression equation:

$$\log\left[I_y(\omega_j)\right] = \beta_0 - d\log\left\{2\sin\left(\frac{\omega_j}{2}\right)\right\}^2 + \vartheta_j,\ j = 1,2,...,g(n), \tag{5}$$

where $\beta_0 = \log f_U(0) + \log f_U(\omega_j)/f_U(0), \vartheta_j = \log I_y(\omega_j)/f_y(\omega_j) - \psi(1), \psi(\cdot)$ being the digamma function, i.e. $\psi(x) = d\log\Gamma(x)$ and $\omega_j = 2\pi j/T$ represent the $m = \sqrt{T}$ Fourier frequencies, $I_y(\omega_j)$ denotes the sample periodogram defined Saganowski:

$$I_y(\omega_j) = \frac{1}{2\pi T}\left[\sum\nolimits_{t=1}^{T} y_t e^{-\omega_j t}\right]^2. \tag{6}$$

The GPH estimator is given by:

$$d_{GPH} = -\frac{\sum_{j=1}^{g(n)}(\nu_j - \hat{\nu})\log I_y(\omega_j)}{\sum_{j=1}^{g(n)}(\nu_j - \hat{\nu})^2}, \tag{7}$$

where $\nu_j = \log\left(2\sin\left(\omega_j/2\right)\right)^2, g(n)$ being the bandwidth in the regression equation. The variance of the GPH estimator is:

$$\mathrm{var}\left(d_{GPH}\right) = \frac{\pi^2}{\sum_{j=1}^{g(n)}(\nu_j - \hat{\nu})^2}. \tag{8}$$

Gewke and Porter-Hudak proved the asymptotic normality of the semiparametric estimator in Equation 7 when $d < 0$ and suggested taking $g(n) = n^{\alpha}, 0 < \alpha < 1$. A detailed description of the presented algorithm can be found in the work of [1].

2.3 The Identification and Estimation of an ARFIMA Model

For the use of the regression techniques several steps are necessary to obtain an ARFIMA model for a set of time series data and these are given below (see [4] and [5] [6]).

Let $\{Y_t\}$ be the process as defined in Section 2.1. Then $U_t = (1-L)^d Y_t$ is an $ARMA(p,q)$ process and $Z_t = \frac{\Phi(L)}{\Theta(L)} Y_t$ is an $ARFIMA(0,d,0)$ process.

Model Building Steps:

1. Estimate d in the $ARIMA(p,d,q)$ model, denote the estimate by $\hat{d}$.
2. Calculate $\hat{U}_t = (1-L)^{\hat{d}} Y_t$.
3. Using Box-Jenkins modeling procedure (see [7] [8]) identify and estimate ϕ and θ parameters in the $ARMA(p,q)$ process $\Phi(L)\hat{U}_t = \Theta(L)\varepsilon_t$.
4. Calculate $\hat{Z}_t = \frac{\hat{\Phi}(L)}{\hat{\Theta}(L)} Y_t$.
5. Estimate d in the $ARFIMA(0,d,0)$ model $(1-L)^{\hat{d}} \hat{Y}_t = \varepsilon_t$. The value of $\hat{d}$ obtained in this step is now the new estimate of d.
6. Repeat steps 2 to 5, until the estimates of the parameters d, ϕ and θ converge.

In this algorithm, to estimate d we use the regression methods described in Section 2.2. It should be noted that usually only one iteration with Steps $1-3$ is used to obtain a model (see, for instance, [5]). Related to Step 3, it has widely been discussed that the bias in the estimator of d can lead to the problem of identifying the short-memory parameters. This issue has been investigated by [9], [10]. For more information about this algorithm you can find in [11].

3 Experimental Results

In this section we compare results achieved for ARFIMA based anomaly detection to SNORT [12] based preprocessor which we proposed in [13]. Preprocessor uses DWT - Discrete Wavelet Transform and Chi-square statistic test for anomaly detection.

Efficiency of ARFIMA based anomaly detection algorithm was evaluated by simulating different attacks on test LAN network. We used Back Track [14] Linux distribution in order to simulate different attacks such Saganowski: Application specific DDos, various port scanning, DoS, DDoS, Syn Flooding, pocket fragmentation, spoofing and others. We used the same set of attacks as in [13] in order to compare ARFIMA based solution to algorithms based on DWT and Chi-square [13].

For algorithms evaluation we extracted 25 features from network traffic (see Table 1). In Tables 2 and 3 there are results of DR detection rates and FP false positive respectively. We can see that for a given test ARFIMA gives us comparable and slightly better results in comparing to DWT and Chi-square based anomaly detection based methods. We can notice that detection rate and false positive strongly depends on given traffic feature. Attack has got direct impact only on selected traffic features from Table 1. f_9 and f_{10} features gives us the best results. $DR[\%]$ for f_9 and f_{10} changes in boundaries $73.68 - 96.52$ in turn $FP[\%]$ changes in boundaries $0.05 - 9.13$.

Table 1. Network traffic features used for experiments

f_1	number of TCP pockets	f_{14}	out TCP pockets (port 80)
f_2	in TCP pockets	f_{15}	in TCP pockets (port 80)
f_3	out TCP pockets	f_{16}	out UDP datagrams (port 53)
f_4	number of TCP pockets in LAN	f_{17}	in UDP datagrams (port 53)
f_5	number of UDP datagrams	f_{18}	out IP traffic [kB/s]
f_6	in UDP datagrams	f_{19}	in IP traffic [kB/s]
f_7	out UDP datagrams	f_{20}	out TCP traffic (port 80) [kB/s]
f_8	number of UDP datagrams in LAN	f_{21}	in TCP traffic (port 80) [kB/s]
f_9	number of ICMP pockets	f_{22}	out UDP traffic [kB/s]
f_{10}	out ICMP pockets	f_{23}	in UDP traffic [kB/s]
f_{11}	in ICMP pockets	f_{24}	out UDP traffic (port 53) [kB/s]
f_{12}	number of ICMP pockets in LAN	f_{25}	in UDP traffic (port 53) [kB/s]
f_{13}	number of TCP pockets with SYN and ACK flags		

Table 2. Detection Rate DR [%] for a given network traffic features

Feature	Chi-sqaure	Mallat	ARFIMA	Feature	Chi-sqaure	Mallat	ARFIMA
f_1	5.26	5.26	6.26	f_{14}	0.00	5.26	12.24
f_2	5.26	10.52	12.24	f_{15}	0.00	10.52	12.24
f_3	0.00	10.52	12.24	f_{16}	0.00	0.00	0.00
f_4	15.78	10.52	12.24	f_{17}	5.26	5.26	6.26
f_5	10.52	10.52	12.24	f_{18}	10.52	10.52	12.24
f_6	0.00	0.00	0.00	f_{19}	5.26	5.26	12.24
f_7	0.00	0.00	0.00	f_{20}	10.52	5.26	6.26
f_8	15.78	31.58	35.64	f_{21}	5.26	10.52	12.24
f_9	94.73	94.73	96.52	f_{22}	0.00	0.00	0.00
f_{10}	73.68	94.73	95.45	f_{23}	0.00	0.00	0.00
f_{11}	0.00	5.26	0.00	f_{24}	0.00	0.00	0.00
f_{12}	68.42	78.95	82.24	f_{25}	5.26	0.00	0.00
f_{13}	10.52	10.52	12.24				

Table 3. False Positive FP [%] for a given network traffic features

Feature	Chi-sqaure	Mallat	ARFIMA	Feature	Chi-sqaure	Mallat	ARFIMA
f_1	4.46	7.43	4.22	f_{14}	3.73	7.48	3.24
f_2	4.07	7.99	4.12	f_{15}	3.91	7.17	3.32
f_3	4.49	7.96	4.15	f_{16}	0.02	0.02	0.02
f_4	4.24	6.06	4.11	f_{17}	0.34	0.39	0.39
f_5	4.57	5.62	3.54	f_{18}	3.90	8.74	3.82
f_6	2.86	4.14	2.23	f_{19}	4.37	8.36	3.26
f_7	5.18	5.33	5.98	f_{20}	3.71	8.50	4.55
f_8	4.20	8.28	4.15	f_{21}	3.81	7.09	3.11
f_9	6.69	9.13	5.05	f_{22}	2.36	3.08	1.60
f_{10}	0.47	0.48	0.48	f_{23}	3.76	3.07	3.42
f_{11}	4.07	12.06	2.56	f_{24}	0.02	0.00	0.00
f_{12}	5.42	4.34	0.05	f_{25}	0.37	0.02	0.02
f_{13}	4.15	7.07	4.14				

4 Conclusion

In this article, the structure of statistical model ARFIMA has been presented. The model has been demonstrated as a tool for anomaly detection in network traffic. For parameters estimation and model identification Geweke & Porter-Hudak and Box-Jenkins modelling procedure was used. As a result, satisfactory statistical estimations of the analysed network traffic signals have been obtained. The ARFIMA model is used for prediction of normal statistical behaviour of the analysed network traffic parameters. The process of anomaly detection involves comparison of parameters of normal behaviour and parameters of the real network traffic. The obtained results clearly demonstrate that the proposed methods are successful at detecting anomalies in time series of network traffic.

References

1. Geweke, J., Porter-Hudak, S.: The Estimation and Application of Long Memory Time Series Models. Journal of Time Series Analysis (4), 221–238 (1983)
2. Rodriguez, A.C., de los Mozos, M.R.: Improving network security through traffic log anomaly detection using time series analysis. In: Herrero, Á., Corchado, E., Redondo, C., Alonso, Á. (eds.) Computational Intelligence in Security for Information Systems 2010. AISC, vol. 85, pp. 125–133. Springer, Heidelberg (2010)
3. Celenk, M., Conley, T., Graham, J., Willis, J.: Anomaly Prediction in Network Traffic Using Adaptive Wiener Filtering and ARMA Modeling. In: SMC IEEE International Conferenceon Systems, Man and Cybernetics, pp. 3548–3553 (2008)
4. Hosking, J.R.M.: Fractional differencing. Biometrika 68, 165–176 (1981)
5. Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods. Springer, New York (1991)
6. Brockwell, P., Davis, R.: Introduction to time series and forecasting. Springer (2002)
7. Box, G.E., Jenkins, M.G.: Time series analysis forecasting and control, 2nd edn. Holden-Day, San Francisco (1976)
8. Box, G., Jenkins, G., Reinsel, G.: Time series analysis. Holden-day, San Francisco (1970)
9. Schmidt, C.M., Tschernig, R.: Identification of Fractional ARIMA Models in the Presence of Long Memory. Paper presented at the FAC Workshop on Economic Time Series Analysis and System Identification, Vienna (July 1993)
10. Crato, N., Ray, B.K.: Model Selection and Forecasting for Long-range Dependent Processes. Journal of Forecasting, 107–125 (1996)
11. Reisen, V.A., Lopes, S.: Some Simulations and Applications of Forecasting Long-Memory Time Series Models. Journal of Statistical Planning and Inference 80(2), 269–287 (1999)
12. SNORT IDS (2014), `http://www.snort.org/`
13. Saganowski, Ł., Goncerzewicz, M., Andrysiak, T.: Anomaly Detection Preprocessor for SNORT IDS System. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 4. AISC, vol. 184, pp. 223–230. Springer, Heidelberg (2013)
14. BackTrack Linux (2014), `http://www.backtrack-linux.org/`
15. Esposito, M., Mazzariello, C., Oliviero, F., Romano, S.P., Sansone, C.: Real Time Detection of Novel Attacks by Means of Data Mining Techniques. In: ICEIS (3), pp. 120–127 (2005)

16. Esposito, M., Mazzariello, C., Oliviero, F., Romano, S.P., Sansone, C.: Evaluating Pattern Recognition Techniques in Intrusion Detection Systems. In: PRIS, pp. 144–153 (2005)
17. 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)
18. Yaacob, A., Tan, I., Chien, S., Tan, H.: Arima based network anomaly detection. In: Second International Conference on Communication Software and Networks, pp. 205–209. IEEE (2010)

Is European Broadband Ready for Smart Grid?

Kartheepan Balachandran and Jens Myrup Pedersen

Department of Electronic Systems, Aalborg University, Denmark
{kba,jens}@es.aau.dk

Abstract. In this short paper we compare the communication requirements for three Smart Grid scenarios with the availability of broadband and mobile communication networks in Europe. We show that only in the most demanding case - where data is collected and transmitted every second - a standard GSM/GPRS connection is not enough. Whereas in the less demanding scenarios it is almost all of the European households that can be covered by a standard broadband technology for use with Smart Grid.

1 Introduction

During the last years, the European agenda on energy politics has increasingly focused on increasing the amount of renewable energies in the European energy system, largely driven by ambitions of decreasing the carbon footprints while at the same time reducing dependability on countries and regions which are considered unstable. However, the production also becomes more difficult to control; power is produced when the wind is blowing or the sun is shining, rather than when the energy is needed.

In response to this, the concept of Smart Grid has been launched. The idea is that the energy consumption can be more flexible, so that the energy consumption can be adjusted according to the production; It can be achieved through e.g. dynamic pricing, or by allowing operators to turn off various equipment on demand. This can be done in both residential and industrial environments; for residential users it could for example make sense to allow the operator to turn off freezers during peak time as long as a certain minimum temperature is kept, while on the other hand they may use additional energy to lower the temperature during periods with a high energy production.

Another example could be that an operator is allowed to control the charging of electrical cars based on some requirements provided by the customer. This could be that the car should be charged by 7 am in the morning, but apart from that the operator can decide when the charging takes place.

There is still no commonly agreed definition of Smart Grid or its requirements. However, it is clear that communication will be needed between devices in the home (Smart Meters) and the relevant operator. At the moment there are different pilot studies being carried out [5,6], where the requirements differ from simply reading and transferring meter readings with time intervals in the ranges of minutes or hours, to almost real-time two-way communication.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_32

In this paper we will explore to what extend the existing wired and wireless broadband infrastructure in the different European countries is able to meet the requirements of the smart grid systems. This is done by comparing smart grid requirements - based on different scenarios - with the coverage of relevant broadband technologies in the European countries. The study can be taken into account when deciding which Smart Grid solutions are feasible in different countries, and/or to what extend the existing communication infrastructure need to be strengthened in order to facilitate the Smart Grid communication.

The remainder of the paper is organised as follows. Section 2 provides the background, and introduces the communication requirements. In Section 3 we present our Analysis and Results. Section 4 concludes the paper by summarizing and discussing the results.

2 Background and Requirements

As described in the introduction, there is not yet a common definition of specification of Smart Grid. The different approaches that are being explored in various projects vary from being highly connected real-time systems, to systems which are based on frequent or not-so-frequent readings of smart meters.

We previously studied the bandwidth requirement for different scenarios of data collection from smart meters in [1]. In that work we defined a smart meter system which was capable of monitoring total power consumption of a house hold and also measure a detailed power consumption of connected appliances. Some of these appliances could be remotely controlled because they were connected to intelligent power plugs. Over all, more data was sent from a smart meter than the usual total power consumption data.

It is not easy to predict what kind of requirement will arise for smart meters so the paper investigated an extreme scenario as well as more commonly discussed scenarios in smart grid. In this paper we will use some of these scenarios to evaluate if Europe is ready for smart grid for what concerns the communication infrastructure.

The scenarios which are used to evaluate if Europe has the technology to facilitate advanced metering infrastructure for smart grid, are based on sampling intervals and on how often the smart meter has to transfer data to an aggregator that collects data from all households and forward it to the utility company. We choose three scenarios which are investigated in [1] and are shown in Table 1.

The extreme scenario is where the smart meter has to transmit data every second and to at least sample data every half a second in order to ensure sufficient

Table 1. Different sample and transfer intervals for the smart meter

Sampling and Transfer rate			
Sample interval	0.5 sec	1 min	5 min
Transfer interval	1 sec	5 min	15 min

data for eg. control purposes. In the other two, less demanding, scenarios the smart meter has to transfer every 5 min and 15 min. of power consumption data, which is also the sampling rate. Furthermore, the table also shows how often the smart meter has to sample power consumption data before it is transferred.

The bandwidth usage depends on the smart meter system and how much data it is transferring. Based on the work in [1], Table 2 shows the average bandwidth required based on the above scenarios.

Table 2. Average bandwidth required at different transfer intervals for 1 smart meter [*bps*]

Average bandwidth required [*bps*]			
	Transfer time		
Sample time	*1 sec*	*5 min*	*15 min*
0.5 sec	10096	1425,8	1405,8
1 min	0	40,37	21,44
5 min	0	31,41	11,96

From the table it can be seen that sampling every half a second and transferring the data at intervals of 1 sec, 5 min or 15 min. gives the highest upstream requirement, over 10000 bps for the extreme case. For the less demanding transfer intervals (5 min. and 15 min.), an average connection of less than 50 bps is required. A GSM/GPRS connection can provide up to 171,23 kbps [4], and therefore from a bandwidth point of view, GSM can facilitate smart grid metering infrastructure. However, from a latency perspective GSM will not be sufficient because it is a shared resource with a higher latency that 1 second.

Based on this previous study, we conclude that in the second and third case a GSM/GPRS connection will be sufficient to cover the communication requirements. However, all standard technologies (DSL, VDSL, FTTP, WiMAX and Cable) as well as HSPA and LTE are able to accommodate the requirements in all cases.

Based on the coverage data obtained from [2] and the data for GSM/GPRS coverage obtained from [3] we are then able to get an overview of how well Europe is covered with communication technologies that can accommodate Smart Grid.

3 Analysis and Results

3.1 Scenario 1

Scenario 1 is the most demanding scenario from a latency perspective and therefore GSM/GPRS coverage is not sufficient. In [2] statistics are provided for coverage of the fixed/standard and mobile technologies separately. In this case standard means the wired connection as e.g cable, dsl etc. For example, for Austria the Total Standard Coverage is 95.0% whereas for HSPA it is 97.5%. Thus, we cannot tell from these data which houses are covered by both Standard Coverage and HSPA. For this study, we have chosen to take the highest of the two

values as an indicator for the total coverage of Standard and HSPA technologies, knowing well that this number might be an underestimate (e.g. Austrian households covered by Standard technologies but not HSPA would not be counted in). We will refer to this value as the aggregated coverage of wired connections and HSPA.

The coverages for the different countries are shown in Fig. 1. It can be seen that the coverage is generally high, i.e. between 90% and 100% in all countries.

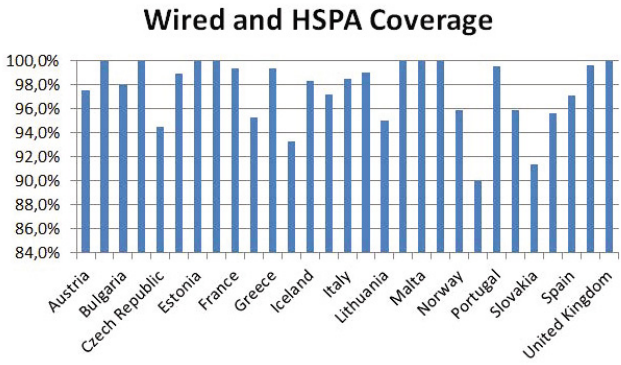


Fig. 1. Aggregated coverage of wired connection and HSPA

3.2 Scenario 2 and 3

We have not been able to find recent reports of European GSM/GPRS coverage. However, [3] from 2006 provides figures for GSM/GPRS coverage throughout Europe back then. In the report, the data are shown graphically so there can be small interpretation errors. It is also clear that the numbers are outdated, which can be seen from the fact that HSPA coverage in some countries are larger than the GSM/GPRS coverage of this report. In any case, we will assume that coverage has not decreased since then, supporting the conclusion that in far most of the countries there is an almost 100% coverage of these technologies. Exceptions with 98-99% coverage are Austria, Bulgaria, Iceland, Italy, Latvia, Norway, Slovakia and Slovenia. For Romania the coverage was around 96%.

4 Conclusion and Discussion

In this paper we wanted to explore whether the mobile and broadband infrastructure in Europe is ready to support Smart Grid. The study we did was based on the communication requirements for three different scenarios for data exchange, compared to European statistics for broadband and mobile coverage. In the most demanding scenario either a fixed connection or at least HSPA would be needed, whereas for the two other scenarios GSM/GPRS was deemed sufficient.

The coverage statistics shows that the coverage of Europe is quite good: Most countries have 95-100% coverage for the most demanding scenario, and either 100% or close to 100% coverage for GSM/GPRS. We believe that these numbers in it self demonstrate that for the vast majority of households in Europe the communication infrastructure can easily support the Smart Grid scenarios. Also, we believe that the benefits of Smart Grid and Smart Meters can be realized even if not each and every household would be connected.

It is worth noting that we have not in this study taken into account that Internet connections are used also for other purposes than Smart Grid communication, so that the aggregated bandwidth requirements may exceed the speeds available. However, since the Smart Grid requirements are relatively small we consider this to be part of a more general discussion on broadband requirements.

Also, for the wireless technologies we have not made any calculations on the usage of shared spectrum, which could in principle be a problem if all users within a cell need to communicate at the same time.

References

1. Balachandran, K., Olsen, R.L., Pedersen, J.M.: Bandwidth analysis of smart meter network infrastructure. In: 2014 16th International Conference on Advanced Communication Technology (ICACT), pp. 928–933 (2014), doi:10.1109/ICACT.2014.6779095
2. European Commision, Coverage Data Tables (2012), `http://ec.europa.eu/digital-agenda/en/fast-and-ultra-fast-internet-access-analysis-and-data`
3. Universal Access. How Mobile can Bring Communication to All (2006), `http://www.gsmworld.com`
4. Mikkonen, J., Turunen, M.: An integrated QoS architecture for GSM networks. In: IEEE 1998 International Conference on Universal Personal Communications, ICUPC 1998, p. 733011 (1998), doi:10.1109/ICUPC.1998.733011
5. Benzi, F., Anglani, N., Bassi, E., Frosini, L.: Electricity Smart Meters Interfacing the Households. IEEE Transactions on Industrial Electronics 5(10), 4487–4494 (2011), doi:10.1109/TIE.2011.2107713
6. Depuru, S., Lingfeng, S.S.R., Devabhaktuni, W., Gudi, V., Smart, N.: meters for power grid Challenges, issues, advantages and status. In: Power Systems Conference and Exposition (PSCE), pp. 1–7. IEEE/PES (2011), doi:10.1109/PSCE.2011.5772451

A Simulator Concept for Cloud Computing Infrastructure

Sławomir Hanczewski[1], Marta Kędzierska[1], and Maciej Piechowiak[2]

[1] Chair of Communication and Computer Networks, Poznan University of Technology, Polanka 3, 60-965 Poznan
slawomir.hanczewski@put.poznan.pl
[2] Institute of Mechanics and Applied Computer Science, Kazimierz Wielki University, Kopernika 1, 85-074 Bydgoszcz

Abstract. This paper presents a concept for a simulator for cloud computing environment. According to the assumptions, and in intention, the simulator will make studies on cloud efficiency (from the cloud operator's perspective) possible. One of its possible applications will be then the evaluation of cloud effectiveness in a situation where, for example, new services have been introduced or existing service parameters have been changed. Results obtained by this simulator can be also used in validation and verification of analytical models proposed for a cloud environment.

1 Introduction

The recent rapid development of a variety network devices and involved technologies that we have been witnessing over the past years has made it possible to re-evaluate the full spectrum of service execution for telecommunications networks [2]. Transmission bitrates offered in modern networks have created a new situation in which the distance from the client to the server or necessary amount of transferred data have ceased to be of any significant importance from the point of view of the quality of the execution of services. The network has ceased to be the bottleneck in telecommunications systems and nothing prevents us now from offering these services from any any place. And this is the essence of cloud computing. On account of the absence of precisely defined standards, the literature of the subject abounds in a number of definitions of the notion of cloud computing. In the present authors' opinion, the best definition so far is the definition proposed by the National Institute of Standards and Technology (NIST) which runs as follows: "Cloud computing is a model for enabling ubiquitous, convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [8].

The concept of cloud computing is undoubtedly an important step in the development of the service market. Its significance is comparable to the introduction of multirate traffic in telecommunications networks. However, the idea

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_33

itself of providing service seems to resemble the idea of mainframe computers that became increasingly popular in the 1970s and the 1980s. These computers offered services (most frequently those related to computational services) in which access to them was provided remotely from dedicated, discless terminals.

Cloud computing enjoys widespread enterprise adoption for it turns out that the cost of a development of an infrastructure of one's own, or a purchase of a licensed software, is more often than not far more expensive than the use of ready-made solutions offered by service providers in cloud. The concept of cloud has an outstanding rank on the priorities stair of governments of individual countries or large international organizations, such as the European Union. In September 2012 the European Commission approved an appropriate strategy for the development of cloud computing in the EU [1]. The strategy outlines the plans for the application of cloud in different domains of life and acknowledges its influence upon the economy, including creation of new work places and the improvement in competitiveness of small and medium-size businesses.

An important problem that still remains to be solved is the efficiency of the infrastructure that performs services in cloud. An addition of new services or, possibly, an increased number of users, can have strong influence upon the efficiency of the infrastructure. How these, and other factors, influence the operation of cloud (i.e., the quality of offered services) can be determined by the simulator whose framework is discussed and proposed in the present paper. This simulator can be also used in verification of analytical models proposed for a cloud (similar like in case models proposed for other telecommunication systems e.g., [9], [5] or [6]) The simulator will be implemented in the C++ language. Currently, there is a number of cloud simulators available. One of them, for example, is CloudSim [3] whose operation is based on jProfiler (Java) [7].

The remaining part of the paper is structured as follows: Chapter 2 outlines the basic assumptions of cloud computing. The concept of the simulator is presented in Chapter 3. Section 4 includes a description of a plan for an implementation of simulator. Section 5 sums up the paper.

2 Cloud Computing

According to the definition proposed by NIST, cloud computing is a model for service provision, understood as access to the infrastructure and/or the application that are available to the user at any randomly selected moment of time. (Fig. 1). According to the assumptions, cloud computing is to be characterized by the following features [8]:

- On-demand self-service - it is the user himself/herself that decides when and from which resources they want to make use of (e.g. the working time of the server); access to required services is then automatic;
- Broad network access - the infrastructure executing cloud computing should be connected with the network via broadband links that make troublefree access to all available services possible. The technologies that make access

to service possible should also support mobile devices such as smart phones or tablets;
- Resource pooling - the infrastructure performing cloud computing is to render service to many users that make use of different services. Available resources should be allocated on demand, according to the demands for a given service. The user has no knowledge and is not aware of where exactly a given service is executed (the user has access to the service no matter where his actual location is).
- Rapid elasticity - resources are allocated and released depending on the current demand thus securing the optimum use of available resources.
- Measured service - The optimum use of available resources is possible thanks to automatic control of executed services. It is also important from the point of view of the user because thanks to constant monitoring of resources used, it is possible to evaluate precisely the costs involved in cloud services.

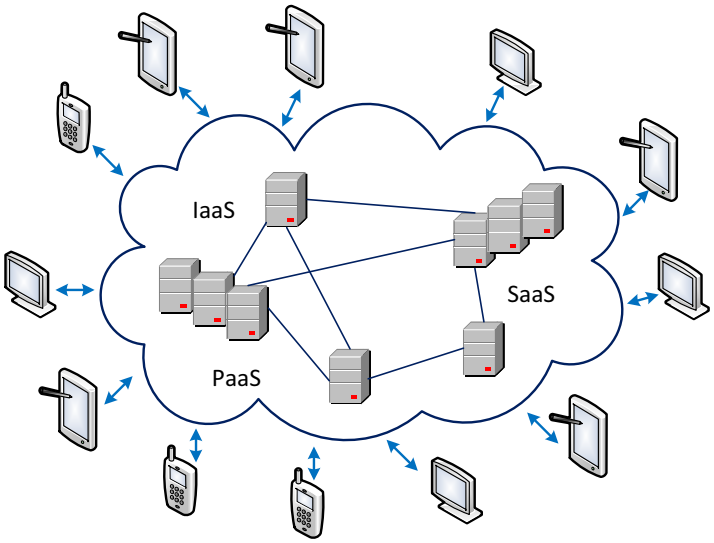


Fig. 1. Idea of cloud computing

Four deployment models in cloud computing are distinguished [8]:

- Private cloud - the cloud infrastructure is dedicated exclusively to a given organization;
- Community cloud - a section of the Private cloud - dedicated to a given specified group of users that require particular solutions (e.g. those stemming from the point of view of security or availability of selected applications);
- Public cloud - a cloud open to all users;
- Hybrid cloud - a cloud that combines some selected features of two or three previously described clouds.

2.1 Cloud Computing Services

Services available in cloud computing have been grouped into three basic categories [8].

The one group of services is composed of the IaaS (Infrastructure as a Service) services. While making use of this particular group of services, the user has access to some parts of the infrastructure of the cloud. Users cannot manage the cloud directly, but can independently decide which software (including the operating system) they are going to make use of. In practice, these services are executed through virtual machines. Charges are calculated depending either on the capabilities of the available virtual machine or depending on its real usage by the user.

The other group of services includes the PaaS (Platform as a Service) services. Within this group of services, the user has access to a selected group of software programs that constitute one platform designed for enabling users to execute their particular tasks. An example of this type of services can be access to the programming environment that would allow its user to create and test various applications. Quite frequently, this type of services is executed using the WWW interface. Also in this case the user has no control over this part of the infrastructure of the cloud that is involved in the execution of a given service and has no influence on the operating system under which the available platform works. Charges for services are calculated according to how much of the resources have been used.

The last group of services includes the SaaS (Software as a Service) services. These services are based on software distribution. The manufacturer of a software makes it available without a necessity of the installation of this particular software locally on the user's hardware. As a result, all issues related to maintenance, updating or technical support for a given application remain on the part of the service provider. Within these services the user has only access to a given application. Final reckoning of this service is performed most frequently in the form of a monthly subscription.

3 The Concept of the Simulator for Cloud Computing

One of the primary uses of the proposed simulator will be the evaluation of the efficiency of the infrastructure executing cloud in variable and changeable work conditions. These changes may result from changes in the parameters for executed services (the number of calls, service time, demanded resources), an introduction of new services or changes in the physical infrastructure of the cloud. Despite the fact that some publications claim that there are virtually no limits in the capabilities of the cloud [4], [1], one should not forget that the capabilities of the physical infrastructure are not limitless simply because it is composed of a limited number of servers. Therefore, our simulator will make it possible to determine such parameters as the level of the usage of the physical infrastructure of the cloud, losses in calls that appear in the process or the service time that in some particular services can be prone to changes depending on the level of

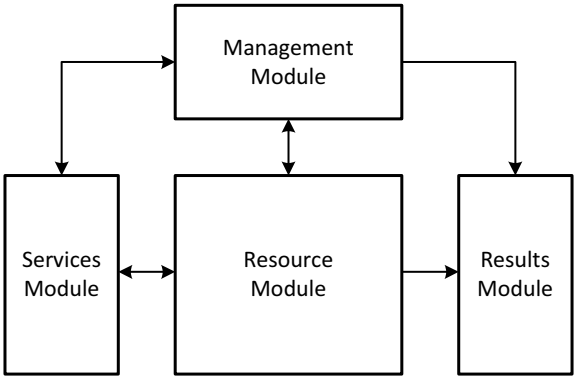


Fig. 2. Flowchart of the proposed simulator

the usage of cloud resources. Fig. 2 shows a flowchart of the proposed simulator. The building blocks of the flowchart include four main modules. The first module (Resource Module) is responsible for the representation of physical resources of the cloud infrastructure. An accurate identification of limited resources of a cloud is absolutely necessary for the accurate operation of the simulator. In general, resources that are used in the execution of services include processing power, disk space, RAM memory and transmission rate. The amount of available data depends on the number of servers, whereas the available transmission bitrate depends on the capacity of the link (or links) that connects the cloud to the Internet. The resource module is assumed to make elastic addition of new servers possible and to represent distributed structures of the cloud (execution of the physical infrastructure composed of two or more data centers).

The cloud offers very diversified services. Moreover, when services are being executed, the degree of the usage of these resources changes in time (e.g. CPU usage and RAM memory usage by the virtual machine). Hence, similarly as in the real cloud, it is necessary to implement the resource Management and new call admission module. This module constantly monitors the usage of resources. In particular cases (depending on the adopted scenario of service execution)it can, for example, limit resources available for one service so that other services can be executed as well. It will also be possible to switch on or off servers at times when appropriate high or low level of resource usage are applicable. Admission of new calls for service will depend on the demands of this particular call as well as on the current level of resource usage. Currently, work on a very precise and accurate determination of the algorithm for the operation of the managing module is being done.

The next module is the Services Module. In line with the assumptions, it is responsible for the description of services. A description of a single service includes the type of the service, demanded resources, average service time, intensity of calls (the number of calls in a time unit) and service scenario. The service scenario includes not only a description as to how the service is executed in the

cloud, but also distributions that describe the usage of cloud resources within the time of its execution (e.g. the RAM memory usage). A determination of occupancy distributions in resources will be the key element that will ultimately influence the accuracy of the operation of the simulator.

The last module to be described is the module responsible for gathering and processing results of the simulation. Technically, appropriately established simulation studies require expected values and confidence interval for measured parameters to be determined on the basis of measurements. At the time of writing this article, however, no set of parameters whose measurement would be possible has yet been precisely established. For the time being, this set includes, among others, the number of losses (of a call or given services that have not been admitted for service), the real service time and the level of usage of resources of the physical structure of the cloud.

4 Implementation Plan

The assumption is that the simulator would be implemented in the C++ language. To streamline the implementation process, work on the simulator has been divided into the following phases:

- Definition of interfaces between the modules of the simulator – a prerequisite for an accurate cooperation between modules of the simulator is to define the method for and the format of transferred data between them, hence the interfaces between the modules are to be defined first and foremost.
- Specification of individual modules of the simulator. Within this phase, the following is to be developed: algorithms for the operation of the managing module, determination of occupancy distributions of resources by services and, for example, of given resources and their influence on the execution of a service. After the completion of this stage all data necessary for the implementation of the simulator will be ready.
- Implementation of the Resource Module.
- Implementation of the Services Module - the implementation of this module, as well as that of the resource module, can be done concurrently and in a parallel manner.
- Implementation of the Management Module - the implementation of the module will commence after the completion of work on the two previous modules because this module is to use data that can be obtained from the two previous modules. This module will be the most advanced module of the simulator.
- Implementation of the Results Module - results will be recorded in files with a format that would make their processing in one of the most popular spreadsheet programs possible.
- Integration and tests of the simulator - this stage will allow individual elements of the simulator to be combined into one working unit. The relevant tests to be carried out will make it possible to evaluate the accuracy of both specifications and implementation of individual modules of the simulator.

At the initial stage of works no graphic interface is planned to be implemented. After the completion of the main implementation work and the testing process of the accuracy of the operation of the simulator, an appropriate graphic interface will be developed. This interface can be also very useful in training and didactic work.

5 Conclusions

This article outlines a general concept of a construction of a simulator for cloud computing. The simulator will be very useful in examining the behavior of the operation of the cloud in variable working conditions. It will make testing of new services possible and will contribute to optimization of those that are already offered. The simulator will also prove to be a useful tool in validation and verification of analytical models for cloud. After the implementation of the graphic interface, the simulator is planned to be used in didactics. Currently, work is being done on the specifications of interfaces and individual modules of the simulator.

References

1. Unleashing the Potential of Cloud Computing in Europe. European Commission, COM (2012) 529 final (2012)
2. Armbrust, M., et al.: A view of cloud computing. Communications of the ACM 53(4), 50–58 (2010)
3. Calheiros, R.N., Ranjan, R., Beloglazov, A., De Rose, C.A., Buyya, R.: CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software: Practice and Experience 41(1), 23–50 (2011)
4. Dillon, T., Wu, C., Chang, E.: Cloud computing: issues and challenges. In: Proceedings of Advanced 24th IEEE International Conference on Information Networking and Applications (AINA), pp. 27–33 (2010)
5. Glabowski, M., Hanczewski, S., Stasiak, M.: Erlang's Ideal Grading in DiffServ modelling. In: Proceedings of 10th IEEE Africon, pp. 1–6 (2011)
6. Hanczewski, S., Stasiak, M.: Modeling of Video on Demand Systems. In: Kwiecień, A., Gaj, P., Stera, P. (eds.) CN 2014. CCIS, vol. 431, pp. 233–242. Springer, Heidelberg (2014)
7. Jprofiler homepage, `http://www.ej-technologies.com/products/jprofiler/overview.html`
8. Mell, P., Grance, T.: The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology, Special Publication 800-146 (2011)
9. Stasiak, M., Hanczewski, S.: Approximation for Multi-service Systems with Reservation by Systems with Limited-Availability. In: Thomas, N., Juiz, C. (eds.) EPEW 2008. LNCS, vol. 5261, pp. 257–267. Springer, Heidelberg (2008)

The New Model of Limited Availability Group

Sławomir Hanczewski and Wojciech Słowik

Chair of Communication and Computer Networks, Poznan University of Technology
slawomir.hanczewski@put.poznan.pl

Abstract. The paper presents a new model of a limited availability group (LAG). The model is based on the Erlang's Ideal Grading (EIG) model. The proposed model is an effect of works on the application of the EIG model in modeling other telecommunication systems. The results obtained with the aid of the new model have been compared with the results of simulations and an earlier model of a limited availability group described in the literature. They confirmed a satisfactory accuracy of the proposed model.

1 Introduction

A limited availability group (LAG) is a model of a system composed of k identical separated subgroups to which a common stream of multirate traffic is offered. The group can only service a call of a given class if it can be serviced by a single subgroup. The LAG model [10] is very often used for modeling other telecommunication systems, e.g., switching networks with unicast and multicast traffic [5,6].

The paper introduces a new model of a limited availability group. The operation of the model is based on the Erlang's Ideal Grading model (EIG) with multirate traffic [4]. The structure of EIG is described with parameters (d, V), i.e., availability and capacity. It should be noted that the availability parameter may be different for each and every class of serviced calls. Numerous studies have shown that by adopting adequate values of availability, one can achieve the same results using EIG as in other telecommunication systems. So far, models were proposed for switching networks with multicast traffic [6], systems with reservation [12], Video on Demand systems [7] or DiffServ network architecture [3]. Continuing these studies, works were conducted in order to develop a proper method of determining the value of the availability parameter for a limited availability group.

The paper is organized as follows: Chapter 2 describes the method of determining the blocking probability in a LAG with multirate traffic, as known from the literature. Chapter 3 presents the model of ErlangâĂŹs Ideal Grading with multirate traffic streams. The proposed new model of a limited availability group is presented in Chapter 4. In Chapter 5, the results of analytical calculations are compared with simulation results. Chapter 6 concludes the paper.

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_34

2 Limited Availability Group

Fig. 1 presents a LAG diagram which is offered m independent streams of calls with the respective intensities: λ_1, λ_2, ..., λ_m. An i-class call demands t_i BBUs for a connection to be established[1]. A limited availability group is described with the following structural parameters (Fig. 1):

- k - the number of subgroups,
- f - the subgroup capacity,
- V - total group capacity ($V = kf$).

LAGs have been subject to many studies. The approximate method of determining the occupancy distribution in a LAG has been proposed in [10]. According to this method, the occupancy distribution in a group may be determined on the basis of a generalized Kaufman-Roberts formula [1], [10]:

$$nP(n) = \sum_{i=1}^{m} a_i t_i \sigma_i (n - t_i) P(n - t_i), \tag{1}$$

where:

- $P(n)$ is the occupancy probability in group n BBU,
- a_i is the traffic offered by calls of the class i,
- t_i is the number of BBUs demanded by the call of the class i,
- $\sigma_i(n)$ is the conditional probability of transition, which may be approximated with the following correlation:

$$\sigma_i(n) = \frac{F(V-n,k,f) - F(V-n,k,t_i-1)}{F(V-n,k,f)}, \tag{2}$$

where $F(x,k,f)$ is the number possible distributions x of free BBUs in k subgroups of which each has a capacity of f BBUs. The value of $F(x,k,f)$ is determined on the basis of the following combinatorial formula:

$$F(x,k,f) = \sum_{i=0}^{\left\lfloor \frac{x}{f+1} \right\rfloor} (-1)^i \binom{k}{i} \binom{x+k-1-i(f+1)}{k-1}. \tag{3}$$

The probability of i-class call blocking in a LAG is determined on the basis of the following formula:

$$E_{i,LAG} = \sum_{n=0}^{V-t_i} P(n)[1 - \sigma_i(n)] + \sum_{n=V-t_i+1}^{V} P(n). \tag{4}$$

[1] in contemporary broadband systems, the smallest unit of capacity is the base bandwidth unit (1 BBU) [8]. It is defined as the greatest common divisor of the value of the resources demanded by all streams of calls offered to the system.

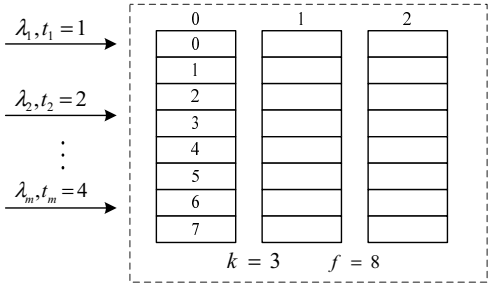


Fig. 1. Limited Availability Group ($k = 3$, $f = 8$, $V = 24$)

3 Erlang Ideal Grading

The fundamental feature of the Erlang Ideal Grading (distinguishing it from other groups) is that the individual sources of traffic do not have access to all V BBUs of the group, but rather to a part of them only [2], [4], [9]. The number of initial BBUs accessible to traffic sources is called availability and marked with the symbol d (Fig 2). The traffic sources with access to the same BBUs of the group form the load group. The number of load groups are marked with g. Moreover, the number of the load groups is always equal to the number of possible ways of selecting d BBUs from all V BBUs (two component groups differ in at least one BBU):

$$g = \binom{V}{d}. \tag{5}$$

In [11], an EIG model with multirate traffic has been introduced. The model also assumed that the availability of all call classes is equal. In [12], a model in which each class could have a different availability has been proposed. It also suggested a calculation method for a case where the availabilities adopted fractional values. According to [12], each class of calls is characterized by a different availability. It means that each class of calls is linked to a different

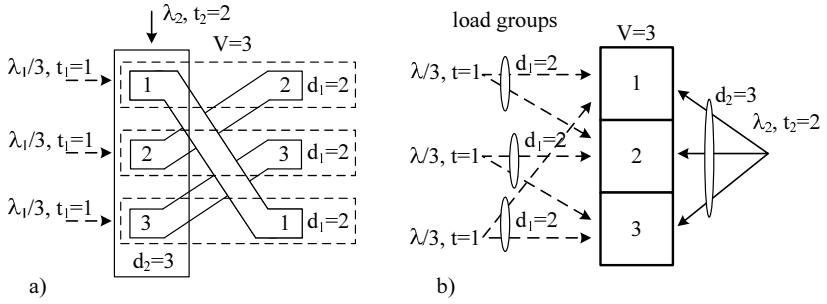


Fig. 2. Erlang's Ideal Grading with multirate traffic and different availabilities [4] (a) concept of availability, (b) offered traffic distribution

number of load groups. Fig. 2 presents a simple diagram of the EIG with the capacity of 3 BBUs. The group manages two classes of calls with the respective availabilities of $d_1 = 2$, $d_2 = 3$. Thus, the number of load groups for the respective classes of calls equals:

$$g_1 = \binom{V}{d_1} = \binom{3}{2} = 3,$$

$$g_2 = \binom{V}{d_2} = \binom{3}{3} = 1.$$

The occupancy distribution in the EIG with multirate traffic and different availabilities may be determined (as in the case of a LAG) on the basis of a generalized Kaufman-Roberts formula (equation (1)):

$$nP(n) = \sum_{i=1}^{m} a_i t_i \sigma_i (n - t_i) P(n - t_i), \tag{6}$$

where the significance of individual parameters is the same as in the case of the LAG. In order to define the conditional probabilities of transition ($\sigma_i(n)$) in an EIG with different availabilities for individual classes of calls, the model proposed in [11] is employed, where the parameters connected with a given class of calls i rely on the availability d_i assigned to this class:

$$\sigma_i(n) = 1 - \gamma_i(n), \tag{7}$$

where $\gamma_i(n)$ is the conditional blocking probability in status n for calls of the class i.

As the occupancy distributions are the same in each component group, the conditional blocking probability for calls of the class i in EIG is equal to the conditional blocking probability in one of the component groups:

$$\gamma_i(n) = \sum_{d_i - t_i + 1}^{k} P_{V,d_i}(n, x), \tag{8}$$

where:

- $k_i = n$, if $(d_i - t_i + 1) \leq (n) < d_i$,
- $k_i = d_i$, if $n \geq d$,
- $P_{V,d_i}(n, x)$ is the probability of the occupancy of x BBUs in a given component group, provided that n BBUs are occupied in the whole group. Probability $P_{V,d_i}(n, x)$ is described with hypergeometric distribution:

$$P_{V,d_i}(n, x) = \binom{d_i}{x} \binom{V - d_i}{n - x} \Big/ \binom{V}{n}. \tag{9}$$

Upon determining the conditional probabilities of transitions ($\sigma_i(n)$), the blocking probability for calls of the class i in EIG servicing multirate traffic may be determined on the basis of the following formula:

$$E_{i,EIG} = \sum_{n=d_i-t_i+1}^{V} P(n)\left[1-\sigma_i(n)\right]. \tag{10}$$

4 A New Model of a Limited Availability Group

The basic characteristic of the proposed method is bringing the calculations of the blocking probability in a LAG (4) to the calculations of the value of such probability in EIG (10), with an identical structure of the offered traffic (A):

$$E_{i,LAG}(A,k,f) = E_{i,EIG}(A,d,V_{EIG}). \tag{11}$$

In order to approximate a LAG through EIG, one needs to specify the values of the structural parameters of EIG, i.e., capacity and availability d_i for individual classes of calls.

The studies conducted by the authors have proved that it is possible to approximate a LAG using EIG, if the capacities of both groups are equal:

$$V_{\mathrm{EIG}} = kf. \tag{12}$$

Adopting the assumption about the equal capacity of both groups has the advantage that the traffic per one BBU in the approximating group is the same as in a LAG. If the identical volume of offered traffic is maintained, the only parameter determining the calculations is the availability for individual classes of calls.

Let us consider the method of determining the availability for any given class i. An i-class call will be serviced by a LAG only if it can be serviced by free BBUs belonging to a single subgroup. Therefore, the number of available BBUs in a single subgroup of this group for an i-class call demanding t_i BBUs, may be formulated as follows:

$$L_i = \left\lfloor \frac{f}{t_i} \right\rfloor t_i. \tag{13}$$

Taking into account the number of subgroups in a LAG, we obtain:

$$d_i = L_i k = \left\lfloor \frac{f}{t_i} \right\rfloor t_i k. \tag{14}$$

Knowing the calculated availability values for all classes of calls, based on formulas (6) – (10), one is able to define the blocking probability in an EIG approximating the LAG.

5 Results

The recommended method of calculating the blocking probability in a LAG is (like the method described in chapter 4) an approximate one. Thus, in order to verify the adopted assumptions, one needs to compare the results obtained using the proposed methods with the results of digital simulation. For this purpose, a LAG simulator has been developed in C++. The simulation employed the event-planning method [13]. Each series simulated 100000 calls of the oldest class (i.e., the one in which the calls demanded the maximum BBU number). It was assumed that the number of series amounted to 5, which allowed for determining a confidence interval of 95%, at least an order of magnitude lower than the value of simulation results. The simulations were conducted for various values of traffic offered per one BBU (parameter a_{BBU}):

$$a_{BBU} = \frac{\sum_{i=1}^{m} a_i t_i}{kf}. \tag{15}$$

Figures 3 and 4 present the results of the comparison of analytical and simulation models. Fig. 3 shows the results for a LAG with the parameters: $V = 48$ BBUs, $k = 3$ i $f = 16$ (System 1). The group served three classes of calls, demanding $t_1 = 3$, $t_2 = 7$ and $t_3 = 13$ BBUs, respectively. Traffics are offered in the following ratios: $A_1t_1 : A_2t_2 : A_3t_3 = 1 : 1 : 1$.

Fig. 4 provides the results for a LAG with the parameters: $V = 24$ BBUs, $k = 3$ i $f = 8$ (System 2). The group served three classes of calls, demanding $t_1 = 1$, $t_2 = 2$ and $t_3 = 3$ BBUs, respectively. Traffics are offered in the following ratios: $A_1t_1 : A_2t_2 : A_3t_3 = 1 : 1 : 1$.

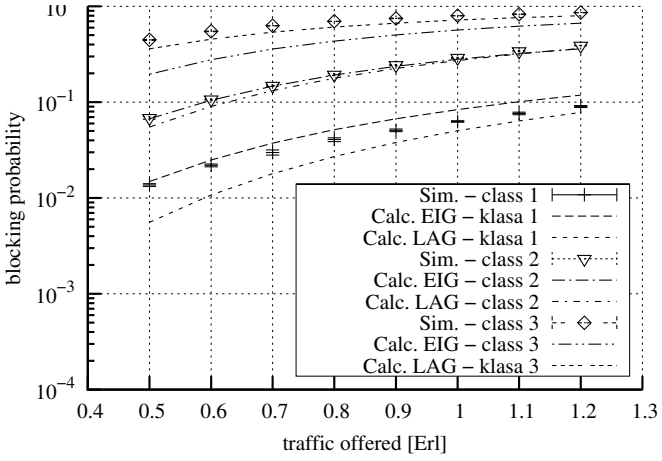


Fig. 3. Blocking probability in Limited Availability Group (System 1: $k = 3$, $f = 16$, $t_1 = 3$, $t_2 = 7$ and $t_3 = 13$)

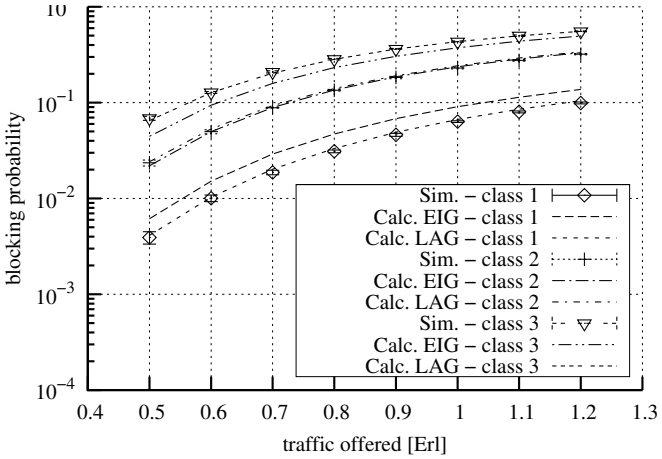


Fig. 4. Blocking probability in Limited Availability Group (System 2: $k = 3$, $f = 8$, $t_1 = 1$, $t_2 = 2$ and $t_3 = 3$)

Figures 3 and 4 present the results of the simulation (Sim.) and the results obtained using the recommended method (Calc. EIG), and the method discussed in chapter 3 (Calc. LAG).

6 Conclusions

The authors of the paper introduced a new method of calculating the blocking probability in a limited availability group with multi-rate traffic. The comparison of analytical calculations and simulation experiments has confirmed the satisfactory accuracy of the proposed method. The calculations conducted according to the method are not complicated. The accuracy of the method may be improved through a more meticulous specification of the availability value. The attained results demonstrate the universal nature of the EIG model which may be used for modeling other telecommunication systems.

References

1. Beshai, M., Manfield, D.: Multichannel services performance of switching networks. In: Proceedings of 12th International Teletraffic Congress, pp. 857–864 (1988)
2. Brockmeyer, E., Halstrøm, L., Jensen, A.: The Life and Works of A. K. Erlang. Transactions of the Danish Academy of Technical Sciences 2 (1948)
3. Glabowski, M., Hanczewski, S., Stasiak, M.: Erlang's Ideal Grading in Diff-Serv modelling. In: Proceedings of 10th IEEE Africon, pp. 1–6 (2011), doi:10.1109/AFRCON.2011.6072139
4. Glabowski, M., Hanczewski, S., Stasiak, M., Weissenberg, J.: Modeling Erlang's Ideal Grading with Multirate BPP Traffic. Mathematical Problems in Engineering 35 (2012), article ID 456910, doi:10.1155/2012/456910

5. Glabowski, M., Sobieraj, M., Stasiak, M.: Modeling Switching Networks with Multi-service Sources and Point-to-group Selection. In: Proceedings of 18th Asia-Pacific Conference on Communications, pp. 686–691 (2012), doi:10.1109/APCC.2012.6388282
6. Hanczewski, S., Stasiak, M.: Point-to-group Blocking in 3-stage Switching Networks with Multicast Traffic Streams. In: Dini, P., Lorenz, P., de Souza, J.N. (eds.) SAPIR 2004. LNCS, vol. 3126, pp. 219–230. Springer, Heidelberg (2004)
7. Hanczewski, S., Stasiak, M.: Modeling of Video on Demand Systems. In: Kwiecień, A., Gaj, P., Stera, P. (eds.) CN 2014. CCIS, vol. 431, pp. 233–242. Springer, Heidelberg (2014)
8. Roberts, J. (ed.): Performance Evaluation and Design of Multiservice Networks, Final Report COST 224. Commission of the European Communities, Brussels (1992)
9. Lotze, A.: History and Development of Grading Theory. In: Proceedings of 5th International Teletraffic Congress, pp. 148–161 (1967)
10. Stasiak, M.: Blocking Probability in a Limited-Availability Group Carrying Mixture of Different Multichannel Traffic Streams. Annales des Télécommunications 48(1-2), 71–76 (1993)
11. Stasiak, M.: An Approximate Model of a Switching Network Carrying Mixture of Different Multichannel Traffic Streams. IEEE Transactions on Communications 41(6), 836–840 (1993)
12. Stasiak, M., Hanczewski, S.: Approximation for Multi-service Systems with Reservation by Systems with Limited-Availability. In: Thomas, N., Juiz, C. (eds.) EPEW 2008. LNCS, vol. 5261, pp. 257–267. Springer, Heidelberg (2008)
13. Tyszer, J.: Object-Oriented Computer Simulation Of Discrete-Event Systems. Kluwer Academic Publishers Group (1999)

Hardening Web Applications against SQL Injection Attacks Using Anomaly Detection Approach

Rafał Kozik[1,2], Michał Choraś[1,2], and Witold Hołubowicz[2,3]

[1] ITTI Ltd., Poznań, Poland
mchoras@itti.com.pl
[2] Institute of Telecommunications, UT&LS Bydgoszcz, Poland
chorasm@utp.edu.pl
[3] Adam Mickiewicz University, UAM, Poznan, Poland
renk@amu.edu.pl

Abstract. In this paper we have presented an algorithm for SQL injection attack detection. The proposed method is based on a linear discriminant analysis algorithm (LDA). However, due to the high dimensionality, the classical LDA does not produce satisfactory results. Therefore, in this paper several extensions have been proposed and evaluated in order to increase the effectiveness of SQL injection attack detection.

1 Introduction

Among all attacks targeting web-servers the SQLIA (SQL Injection Attack) still remains one of the most important network threat. It is ranked as one of the top threats in the OWASP list [2]. The code injection and other similar exploits are the results of interfacing a scripting language by directly passing information through another language and are ultimately caused by insufficient input validation. Particularly, the SQL Injection Attacks refer to a code-injection attacks category in which part of the user's input is interpreted as SQL code. Such code, if executed on the database, may change, erase, or expose sensitive data stored in the database.

According to [3] the SQL injection attack types executed with a web-browser can be categorised (with respect to the technique they use) into following groups:

- Blind SQL Injection
  These techniques prepare injection pattern on the basis of the application response (e.g. error messages). The common scenario is that attacker sends series of requests to the database server via web application (e.g. tampering with HTTP request) and examines the results for the injection point.
- Poorly Filtered Strings
  These techniques exploit flaws in a validation mechanism (e.g. inputs are not checked for escape characters).

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_35

- Filter Bypassing
  These techniques adapt different mechanisms in order to bypass user input validation algorithms. For instance, filters such as addslashes() and magic_quotes_gpc can be bypassed when the vulnerable SQL server is using certain character sets such as the GBK character set [10].
- Incorrect Type Handling
  The attacks based on flaw that occurs whenever user data is not checked for its type (e.g. numbers, text, dates).
- Signature Evasion
  Techniques that use different types of encoding (e.g. URL encoding) in order to avoid detection by IPS, IDS systems or validation mechanisms.

There are several tools and methods that are dedicated to combat SQL injection attacks exploiting mentioned above application flaws. Some of the frequently used tools are based on the static code analysis approaches in order to find the vulnerabilities that may be exploited by any cyber attack. Some examples of such tools include PhpMiner II [13], STRANGER [12], AMNESIA [11]. However, as it is stated in [4], the difficulty relates to the fact that many kinds of security vulnerabilities are hard to be found automatically (e.g. access control issues, authentication problems). Therefore, currently such tools are able to automatically find only relatively small fraction of application security flaws.

There are also solutions that adapt signature-based approach to describe (and detect) cyber attacks. Some examples include PHP-IDS [5], SCALP [6], Snort [8]. The biggest advantage of such tools is their ability to process huge amounts of data. This is due to the fact, that there are efficient algorithms that are able to check given piece of text against pattern (usually expressed as PCRE [9] regular expressions) in a short time. However, the common drawback is that an expert knowledge is required to build such patters describing cyber attack. Moreover, such attack like SQL injection are easy to obfuscate (e.g. using URL encoding). Therefore the problem of providing reliable pattern of an attack is difficult to be solve.

This paper is structured as follows. Firstly, we present our approach to feature extraction from HTTP dumps. Then, we give an overview of the proposed algorithms adapted for classification problem of extracted feature vectors. The experiments setup, results and conclusions are provided after.

2 Feature Extraction

The features describing the normal behaviour of an application are extracted from the log files generated by HTTP server. The example of the HTTP log file is shown in Fig. 1. Each single line in the log file represents single request sent by a client to the web server. It contains information about the source IP address the request was sent from, the exact time stamp, the HTTP request method (GET, POST, PUT, etc.), the HTTP protocol version (1.0 or 1.1), and HTTP result code returned by the web server (e.g. 404 when resource was not found on the server).

In this approach, we are extracting only this part of log line, which contains the HTTP request (method, url address and request parameters). As it is shown in Fig. 1, each HTTP request can be of different length. Therefore, they can not be used directly to build a model or learn a classifier. In order to achieve feature vectors of a constant length we have adapted an approach based on histograms. Such method is commonly used for text analysis and produces histograms of letter distribution. In this approach we adapt histograms that length is 256 (one histogram bin per character for ACII encoding scheme).

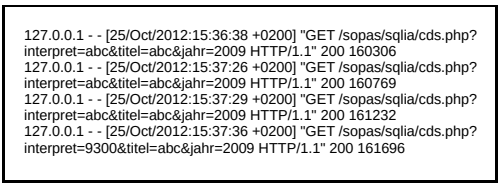


Fig. 1. Example of HTTP log file

3 Adapted Algorithms Overview

Once the feature vectors have been extracted, we apply different machine learning schemas to investigate their effectiveness. In this research we consider two class recognition problem. Therefore, the extracted feature vectors are labelled either as normal or anomalous. In this work we have selected linear discriminant analysis (LDA) to solve the classification problem. However, our experiments showed that its performance is not satisfactory. Therefore, we have introduced several modifications that have been described in the following sections. The modifications include dimensionality reduction (presented in section (3.2)) and adaptation of Simulated Annealing for LDA projection vector computation.

3.1 LDA - Linear Discriminant Analysis

The Linear Discriminant Analysis goal is to reduce dimensionality while preserving the discriminatory information. Assuming that there are two sets of D-dimensional samples x belonging to different classes (normal and anomalous samples), we seek for a projection vector w that will produce scalar y (see eq.1) that will maximize separability of these scalars.

$$y = w^T \cdot x \tag{1}$$

In order to find such a good projection vector, a measure of separation is defined. Commonly Fisher approach (see eq.2) is used, which aims at maximizing the difference between the means $(\mu_1 - \mu_2)$, normalized by a sum of variances $(s_1^2 + s_2^2)$ computed for each class (called also a measure of the within-class scatter).

$$J(w) = \frac{|\mu_1 - \mu_2|^2}{s_1^2 + s_2^2} \tag{2}$$

Fisher criterion is also expressed using S_B and S_W matrices (see eq. 3), where $S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$ and $S_W = \sum_{i \in \{1,2\}} (x - \mu_i)(x - \mu_i)^T$

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{3}$$

The optimal w^* is found solving the generalized eigenvalue problem (eq.4).

$$w^* = \arg\max_w [J(w)] = S_W^{-1}(\mu_1 - \mu_2) \tag{4}$$

3.2 Modified LDA

The high dimensionality of x samples may cause the variances $s_{1,2}$ to be computed inaccurately. Therefore, the the inverse of S_W matrix may be impossible to be computed or the result would carry significant error. In order to solve that problem a dimensionality reduction step may be added prior the LDA. In this work, we have investigated two approaches. First approach computes histogram for each feature vector. Since the feature vector is a histogram itself, this operation is adequate to extending bins width. The second approach adapts SVD (Singular Value Decomposition, see eq.5). The matrix with sample vector (A) is decomposed to left singular vectors matrix (U), singular values matrix (S), and right singular vectors matrix (V).

$$SVD(A) = USV^T \tag{5}$$

In order to reduce the dimensionality, the smallest singular values of S matrix are set to 0 (also corresponding rows of V and S can be eliminated) and the matrix A is composed back.

3.3 Simulated Annealing Discriminant Analysis

The Simulated Annealing (SA) is a generic algorithm for global optimization. In this work we have adapted SA for computing the optimal w^* vector. It is an iterative algorithm, which evaluates single projection vector w which is used to compute the scalar value with vector coming from samples sets (using eq.1). The algorithm provides schema for searching a better solution within a controlled neighborhood. The SA algorithm successively repeats stages of annealing and cooling steps.

Before the algorithm is started, the initial temperature T_0 is set. The SA algorithm proceeds in successive stages of annealing and cooling steps:

– During the "annealing" process, a random initial guess of w^* is chosen in the neighborhood of w^*. The picked solution is accepted (and the old solution w is replaced) with probability $p(w^*)$ described by eq.6, where $E(w^*)$ indicates how efficient is the solution. In our case the efficiency measure is a separability of two classes (normal and anomalous feature vectors).

$$p(w^*) = \exp(\frac{E(w) - E(w^*)}{T_0}) \tag{6}$$

– During "cooling" phase the temperature is decreased to a new temperature $T_1 < T_0$. Then, a new iteration of annealing with the updated temperature T_1 is performed.

The annealing and cooling steps are iterated until the system reaches a global steady state or the cooling temperature reaches certain value.

4 Experiments

4.1 Scenarios Preparation

In this section our evaluation methodology is described. The SQL Injection Attacks are conducted on php-based web service with state of the art tools for services penetration and SQL injection. The traffic generated by attacking tools is combined together with normal traffic in order to estimate the effectiveness of the proposed methods.

The web service used for penetration test is so called XAMPP (Apache + MySQL + PHP) server with MySQL back-end. It is one of the most common worldwide used servers and therefore it was used for validation purposes. The server was deployed on Linux Ubuntu operation system. For penetration tests, sample services developed in PHP scripts and shipped by default with the server are validated.

The reason why we have adapted this server for pen-testing is the fact that multiple SQL injection vulnerabilities have been identified for several default applications deployed on that server (e.g. CD Collection, Biorythm, Guest/Phone Book). Some examples of XAMPP server vulnerabilities extracted from NVD (National Vulnerability Database) database [7] have been presented in Tab.1.

Table 1. Examples of XAMPP server vulnerabilities (source: nvd.nist.gov)

CVE	Description
CVE-2007-2080	Multiple SQL injection vulnerabilities in XAMPP 1.6.0a for Windows allow remote attackers to execute arbitrary SQL commands via unspecified vectors in certain test scripts.
CVE-2005-1077	Multiple cross-site scripting (XSS) vulnerabilities in XAMPP 1.4.x allow remote attackers to inject arbitrary web script or HTML via (1) cds.php, (2) Guestbook-EN.pl, or (3) phonebook.php.
CVE-2008-3569	Multiple cross-site scripting (XSS) vulnerabilities in XAMPP 1.6.7, when register_globals is enabled, allow remote attackers to inject arbitrary web script or HTML via the text parameter to (1) iart.php and (2) ming.php

For evaluation purposes it was necessary to collect both genuine and malicious requests (containing SQL injection attacks) sent to web application.

The genuine traffic was generated using web crawler that was programmed to populate in an automatic way (via the HTTP protocol) the database of pen-tested applications. The crawler also browsed the web pages and sub-pages and perform searching the same way the human user does. The data used to populate the content of the database was as much realistic as possible. In example, for the "CD collection" web application the data was obtained from real CDs collections. Therefore, the attributes stored in the tables (e.g. CD's release date) were also within realistic ranges.

Attack injection methodology was based on the known SQL injection methods, namely: boolean-based blind, time-based blind, error-based, UNION query and stacked queries. For that purpose sqlmap [1] tool was used. It is an open source penetration and testing tool that allows the user to automate the process of validating the tested services against the SQL injection flaws.

While the crawler was populating the database content and performing various other operations (browsing, searching, deleting, etc.) the traffic volume was intercepted. In this scenario two types of data were collected (Fig. 2), namely: HTTP traffic (in form of WWW server log file), and SQL Queries (in form of SQL database log file).

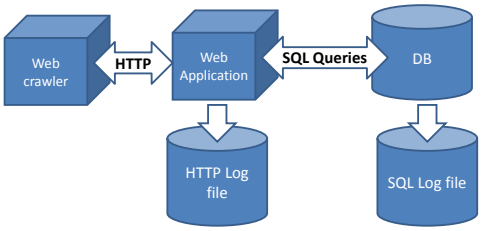


Fig. 2. Testbed configuration overview

4.2 Effectiveness Evaluation Methodology

For evaluation purposes we have adapted the stratified 10-fold cross-validation technique. It is one of the common and standard approaches of predicting the error rates of a classifier that has been established using machine learning algorithms.

For that approach the data obtained for learning and evaluation purposes is divided randomly into 10 parts (sets). For each part it is intended to preserve the proportions of labels (e.g. number of anomalies and normal feature vectors) in the full dataset. One part (10% of full dataset) is used for evaluation while the remaining 90% is used for training (e.g. establishing model parameters). When the classifier is learnt, the evaluation data set is used to calculate the error rates. The whole procedure is repeated 10 times, so each time different part is used for evaluation and different part of data set is used for training. Each single run of evaluation is called a fold. The result for all 10-folds are averaged to yield an overall errors estimates. Despite the fact that there is an open debate

on the best scheme for evaluation in data mining community, the 10-fold cross-validation approach has been accepted as a one of the standard method yielding reliable errors estimations.

5 Results

The results show that classical LDA algorithm (see Tab.2) does not produce satisfactory results. It is a typical problem that for high dimensionality the estimation of variance matrices of analysed feature vectors becomes inaccurate.

The experiments for the first technique of the dimensionality reduction (LDA + histogram) showed that the attack recognition ratio can be increased to 90% (from 60%). The dimensionality reduction based on SVD (LDA + SVD) produces lower detection ratio (86%), but also reduces the false positives to 5%(from 10%). It was noticed that combining both SDA + histogram approaches does not improve results significantly.

The best results are achieved for Simulated Annealing approach. This method allowed us to achieve 98% of detection ratio, while having 6.5% of false positives.

Table 2. Method evaluation results for HTTP log file

Method	Detection Rate	False Positive Rate
LDA	59.94%	29.58%
LDA + histogram	89.94%	10.06%
LDA + SVD	86.94%	**5.0%**
LDA + histogram + SVD	89.77%	9.76%
Simulated Annealing	**98.38%**	6.51%

6 Conclusions

In this paper we have presented an algorithm for SQL injection attack detection. First, we have explained how feature vectors can be extracted from HTTP requests. Then, we have investigated different approaches to establish reliable attack detection method. Due to the high dimensionality,, the classical linear discriminant analysis (LDA) does not produce satisfactory results. Therefore, several extensions have been proposed. The evaluation results proved that estimating LDA transformation vector with Simulated Annealing approach can increase the effectiveness of SQL injection attack detection.

Acknowledgment. This work was partially supported by the Applied Research Programme (PBS) of the National Centre for Research and Development (NCBR) funds allocated for Research Project number PBS1/A3/14/2012 (SECOR).

References

1. Sqlmap project homepage, http://sqlmap.org/
2. OWASP Top 10 – 2013, https://www.owasp.org/index.php/Top_10_2013-Top_10
3. Hakipedia project homepage, http://hakipedia.com/index.php/SQL_Injection
4. Source Code Analysis Tools, https://www.owasp.org/index.php/Source_Code_Analysis_Tools
5. PHP-IDS project homepage, https://phpids.org/
6. Apache Scalp Project homepage, http://code.google.com/p/apache-scalp/
7. National Vulnerability Database homepage, http://nvd.nist.gov/
8. Snort project homepage, http://www.snort.org/
9. Perl-compatible regular expressions (pcre), http://www.pcre.org
10. GBK char set, http://en.wikipedia.org/wiki/GBK#Encoding
11. Halfond, W.G., Orso, A.: AMNESIA: analysis and monitoring for NEutralizing SQL-injection attacks. In: Proceedings of the 20th IEEE/ACM International Conference on Automated Software Engineering, pp. 174–183. ACM (2005)
12. Yu, F., Alkhalaf, M., Bultan, T.: Stranger: An automata-based string analysis tool for PHP. In: Esparza, J., Majumdar, R. (eds.) TACAS 2010. LNCS, vol. 6015, pp. 154–157. Springer, Heidelberg (2010)
13. Shar, L.K., Tan, H.B.K.: Predicting common web application vulnerabilities from input validation and sanitization code patterns. In: 2012 Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 310–313. IEEE (2012)

Versatile Remote Access Environment for Computer Networking Laboratory

Karol Kuczyński[1], Rafał Stęgierski[1],
Waldemar Suszyński[1], and Michael Pellerin[2]

[1] Maria Curie-Skłodowska University, Institute of Computer Science
ul. Akademicka 9, 20-033 Lublin, Poland
karol.kuczynski@umcs.pl
[2] Extreme Networks, 9 Northeastern Boulevard, Salem, NH 03079, USA
pellerin@extremenetworks.com

Abstract. In this paper the integrated remote access system for a computer networking laboratory is presented. It can be used for scientific, training or engineering purposes. Many features of the proposed solutions are superior to commercially available products.

1 Introduction

Advanced systems for integrated network management are commonly known and commercially available [1,2,3]. Computer network equipment manufacturers are focused on solutions for production networks, but little attention is paid to networking laboratories, where networking solutions are used as a study subject rather than enterprise infrastructure. Such laboratories are used mainly for:

- numerous scientific purposes,
- teaching and learning computer networking [4],
- and as a test environment for networking engineers.

In each case there are special needs that are difficult to address with ready-made systems. That is why in most of networking laboratories no special management system is used or their use is limited. Users have direct, physical access to the hardware (routers, switches, controllers, etc.), so that they can freely connect intermediary and end devices, using all ports (including console and power outlets).

This approach seems to be optimal for teaching networking basics only. Beginner students can concentrate on actual devices and cabling, without being distracted by any additional higher level mechanisms and features. For all other user categories, physical access to devices is necessary only to perform cabling, hardware repair or hardware hacking. Then, staying in close neighbourhood of the devices is inconvenient due to heat and noise generated by them, and possibly many other factors. Thus, remote access would be preferred for both users staying in the next room and users in another part of the world (teleworkers or telecommuters [5]).

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_36

A remote access system for networking laboratory is expected to meet the following criteria:

- It should be possible to remotely turn on and off the devices (hard reset happens to be necessary for example during password recovery). Lab devices do not need to operate continuously and should be off when not used, to save energy.
- Console ports (RS-232) of all devices should be accessible simultaneously (typical PC usually has only one or two available RS-232 interfaces). This is often the main out-of-band configuration access method of networking devices.
- Out-of-band access to command line and/or graphical desktop of physical and virtual end devices connected to the laboratory network should be provided.
- Multi-vendor environment is to be supported.
- If virtual PCs are used, it should be possible to assign physical Ethernet or 802.11 network cards to them, to reproduce real environment conditions.
- User rights need to be precisely configured and controlled in a multi-user environment. Wireless devices may need to be isolated from each other. It should be possible to easily (with one click) save device configuration, or create a snapshot of virtual machines and to restore it later. Snapshots and configuration management is an important issue.
- The system is to be highly configurable and customisable.
- Consistent graphical user interface (preferably web-based) should be available for both administration and regular use. Beginner users prefer to use a web browser only and do not like to install additional software.
- Various security issues need to be addressed.
- The system cost is expected to be adequate to offered benefits, with possibly simple licensing model.

The above desired characteristics results from authors' experience of teaching, networking labs preparation and scientific work. Successful implementation of such a system will result with more comfortable working conditions for its users and administrators, and significant savings (costs of travel and laboratory maintaining, more effective and efficient use of lab equipment, reduced "carbon footprint", etc. [5]).

In authors' opinion, none of currently available remote access solutions meet all or at least most of the above criteria. The purpose of the presented work is to design, implement and test such a system. Its main building blocks have been already created and routinely work in the networking laboratory in the Institute of Computer Science of Maria Curie-Skłodowska University. Current works are concentrated on their integration, user management subsystem and unified interface.

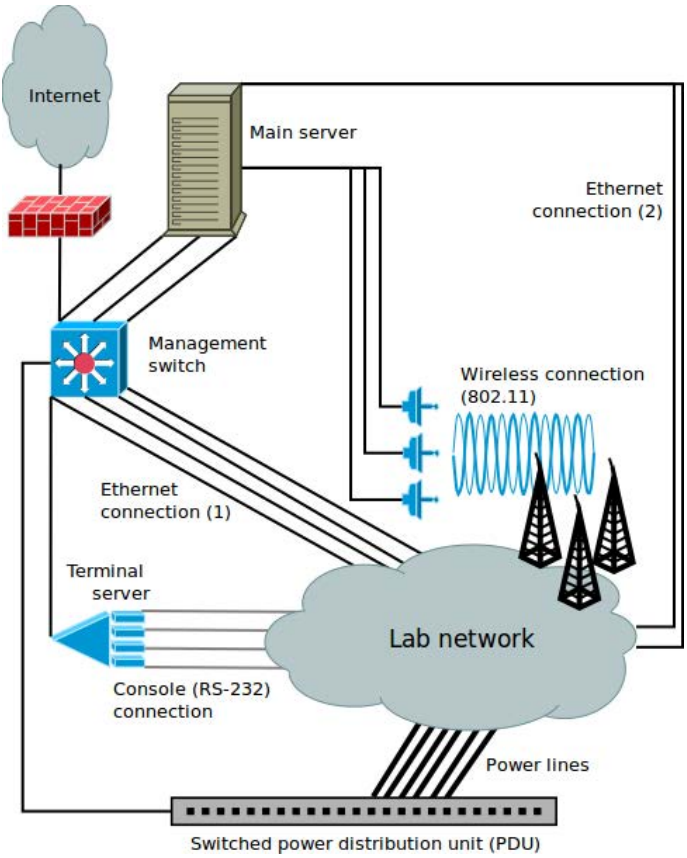


Fig. 1. General topology of the proposed environment

2 Materials and Methods

2.1 Hardware Components

The general topology of the proposed environment is presented in Fig. 1. The main server running under VMware®ESXi™(either free or paid version, depending on requirements) is a heart of the system. It may by equipped with wireless 802.11 network cards, if wireless connectivity between the lab network and virtual PCs running on the server is needed. Virtual machines have virtual Ethernet cards, connected to the virtual switch (IEEE 802.1g [6]) that has a connection to the management switch. Physical Ethernet cards can be also mapped to virtual machines, if necessary (for example for 802.1x authentication on the laboratory network switches).

The terminal server provides network access to console ports of network devices to be managed. It can be a separate device or a multi-port RS-232 PCI card, attached to the server.

The switched power distribution unit makes it possible to remotely (through console, telnet, web interface or SNMP) power on and off the devices.

We assume that devices in the laboratory network are interconnected manually by a local operator. However, if a remotely controlled Layer 1 switch was used, the cabling task could also be performed automatically, according to requirements.

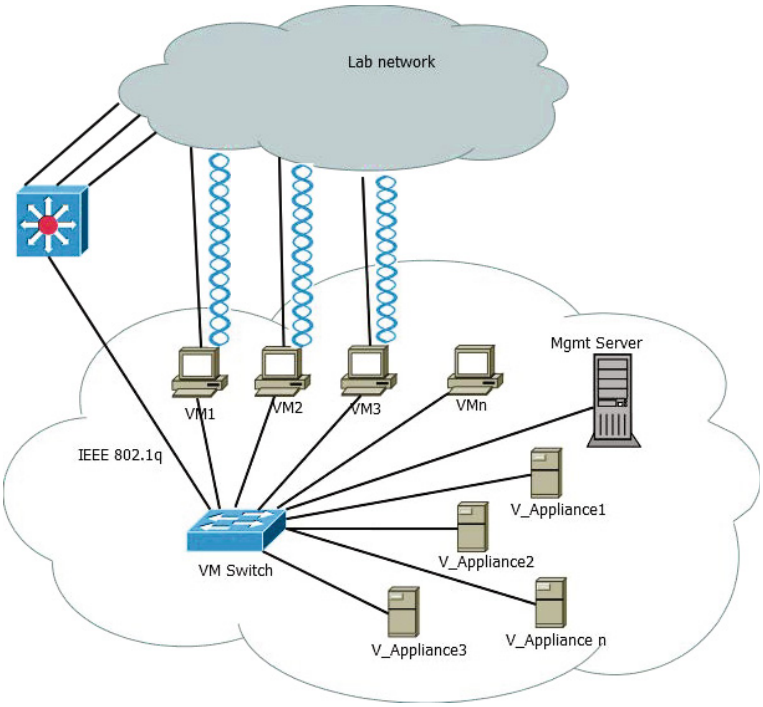


Fig. 2. The virtual part of the environment

2.2 Software Components

The main server (Fig. 1) is used to run both the access environment management software (virtual Mgmt server) and virtual machines (PCs, servers, virtual appliances, etc.) for the laboratory network – Fig. 2. Virtual and physical devices are grouped into bundles (Fig. 3). Users are authorised to access only the bundles assigned to them by the administrator. Additionally, a bundle is visible for a user in slot time reserved in advance, because some bundles are available only in defined periods. It is the result of the fact that hardware could be bonded in different bundles, according to current hardware configuration.

Users can connect to the devices using their own telnet/SSH client, VNC client or through integrated web-based interface (compatible also with mobile devices). The devices are automatically powered on and off. Users can perform a hard reset.

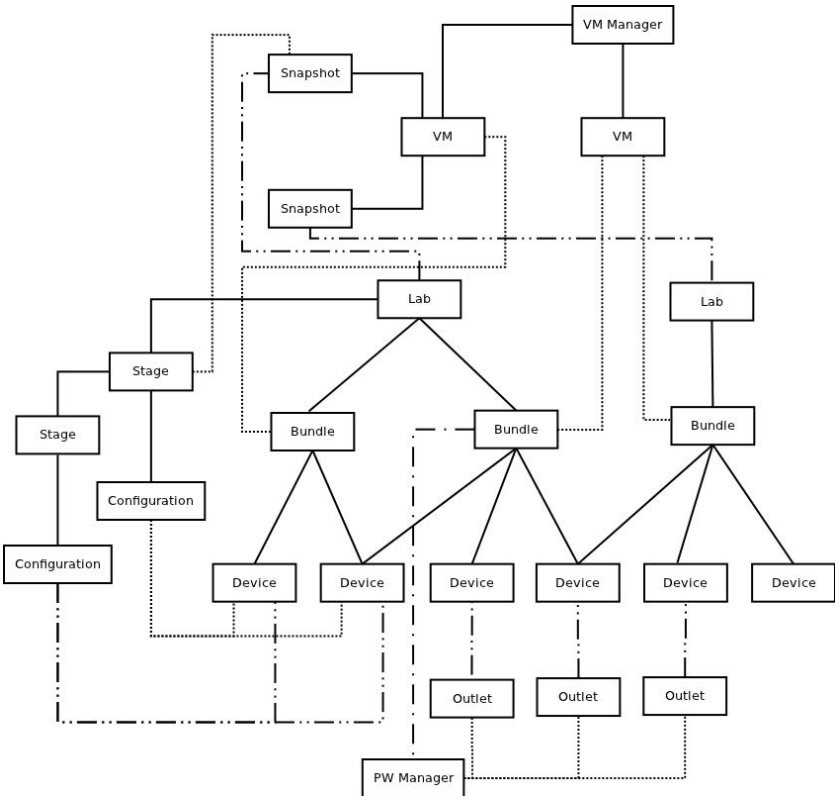


Fig. 3. Management software architecture diagram

All configuration files and virtual machines' snapshots are centrally managed and stored. User's session can be automatically saved and restored later. VLAN configuration of the management switch (Fig. 1) and the VM switch (Fig. 2) is also performed automatically for the bundles that are to be used.

Each laboratory could have stages (Fig. 3) defined by administrator with proper configuration and/or VM snapshot. Such a stage could be restored by a user who wants to have faultless environment for a next part of an exercise.

Web access is build with Symfony framework with Twig templates and Doctrine as O/RM database access engine. Application is based on MVC software pattern. PHP code execution is accelerated with APC. All devices could have defined multiple access types both on and out-of-band. Initial configuration of each device is performed via SNMP communication and/or Telnet/SSH parsed access. Build in VNC and Telnet/SSH client are Java applets and could be, in some cases, configured with tunnelled connection to restrict Web Service to act as a one and only public address of whole laboratory.

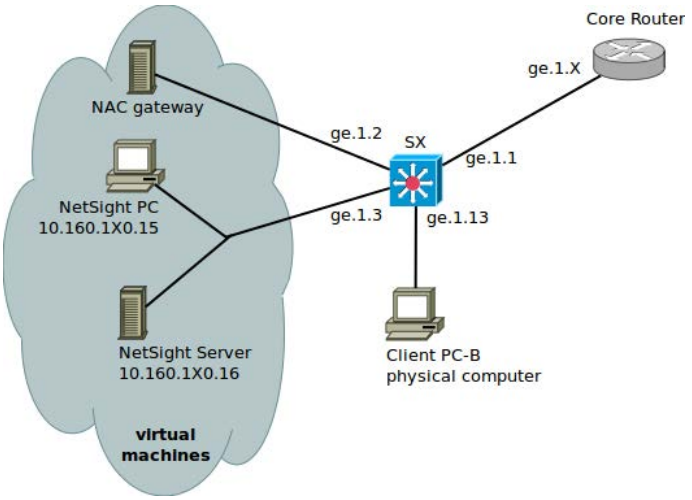


Fig. 4. Example laboratory network topology

3 Results

The presented environment has been lately used for providing Network Access Control (Extreme Networks NAC[1]) course for six teams of students (two students in each team). Each team needs a multilayer switch, two virtual appliances (NetSight appliance and a NAC gateway), two PCs (an administrator's PC and user's PC) and an advanced multilayer switch – Fig. 4. The course was provided in a mixed environment, with virtual machines (NetSight administrator PC, NAC and NetSight) and physical devices (the switch and the user's PC). The main server (Fig. 1) has two Intel Xeon E5606 processors and 36GB RAM. During the classes, 80% of the RAM was used, processor utilisation was between 30 and 50%, and power consumption was around 150W. No problems with performance or access to resources were reported.

The system is also routinely used for advanced wireless LANs courses. Each bundle consists of a wireless controller (either physical or virtual; access to command line and GUI is provided), two wireless access points, and a virtual PC with a physical WiFi card (located within access points' range). A Radius and DHCP server is also accessible.

4 Discussion

Use of commercially available remote access systems for networking laboratories seems to be a good choice when a laboratory is used for regular classes only.

[1] `http://www.extremenetworks.com/product/network-access-control/`

NDG NETLAB+[2] is probably the most well-known environment for information technology courses, but it has some significant limitations. Situation becomes problematic when the same networking infrastructure is used for different courses and scientific purposes, by many users, concurrently. Scientific needs are often non-standard, unpredictable and incompatible with the license agreements. Desired modifications are usually problematic or impossible due to technical or legal issues. The presented environment is free from such problems. It can be freely modified and extended, according to actual needs. Its main elements are based on free license solutions (including VMware ESXi).

The device configuration management subsystem is one of the crucial components. Configuration files of laboratory network devices (also in a multi-vendor environment) and snapshots of virtual machines can be saved and restored any time. Different users can use the same equipment alternately and continue their tasks. It is also easy to offer troubleshooting labs, where a network is preconfigured and students are required to detect and remove mistakes.

Another interesting feature is the choice of access method. The devices can be accessed using either the integrated web-based GUI (this method is preferred by beginners) or user's preferred applications (VNC client, SSH client, SNMP tools).

5 Conclusion

The proposed system has been already implemented and tested. The networking laboratory can be now used either locally or remotely for students' training, scientific and engineering purposes, concurrently. Some features still need to be added or refined, but laboratory management is now easier and less time consuming than before the system implementation, while the laboratory network offers new possibilities. This solution is cost effective, since its main building blocks are based on free license software. The integration software has been designed by the authors. It is modular and easy to modify and expand, according to any future needs.

References

1. Cisco Systems, Cisco Prime Infrastructure 2.0 Data Sheet (2014), http://www.cisco.com/c/en/us/products/collateral/cloud-systems-management/prime-infrastructure/data_sheet_c78-729088.pdf
2. Extreme Networks, NetSight Data Sheet (2014), http://www.extremenetworks.com/product/netsight/
3. Hewlett-Packard Development Company, Network Management (2014), http://www8.hp.com/us/en/software-solutions/software.html?compURI=1171412
4. Bastidas, C.E.C.: 41st Enabling Remote Access to Computer Networking Laboratories for Distance Education. In: ASEE/IEEE Frontiers in Education Conference, Rapid City, SD (2011)

[2] http://www.netdevgroup.com/products/

5. Ellison, N.B.: Telework and Social Change: how technology is reshaping the boundaries between home and work, p. 18. Praeger, Westport (2004) ISBN 9780313051715, OCLC 57435712
6. IEEE 802.1G – Remote MAC Bridging (2006), http://www.ieee802.org/1/pages/802.1G.html

A Low Bitrate Video Transmission for IP Networks

Piotr Lech

West Pomeranian University of Technology, Szczecin
Department of Signal Processing and Multimedia Engineering
26. Kwietnia 10, 71-126 Szczecin, Poland
piotr.lech@zut.edu.pl

Abstract. This article presents the concept of the algorithm for low bitrate network video transmission based on the Monte Carlo image analysis method. Each video frame after dividing into squares is used for collecting data about energy, motion and image histogram. The motion detection is used to decide which of prepared data should be transmitted. The compression ratio of each block can be changed adaptively depending on the value of energy in the square block. Conducted simulations confirmed the algorithm's efficiency for data transmission for low bitrate networks. The IP (Internet Protocol) and GPRS (General Packet Radio Service) networks have been used for real tests. The TFTP (Trivial File Transfer Protocol) has been applied for the test application.

1 Introduction

In this paper an efficient method of image processing based on the Monte Carlo method is presented. The main advantages of the proposed algorithm are low computational cost and possibility of estimation of some motion parameters such as direction and velocity. Practical applications of the proposed algorithm in automatic control systems can be related to inspection and security applications. For this purpose it is possible to utilize some statistical methods, in particular motion detection based on the Monte Carlo method. Using the information from the motion detector corresponding to the local changes on the image (motion) one may decide which fragments of the image are the most important and should be transmitted immediately because of the rapid changes of their contents. It is worth noticing that no information from the visualization part is transmitted back to the acquisition part of the system so the proper direct correction in the acquisition part is not possible. Additionally it should be taken into account that the acquisition side of the system does not have information about any blocks corresponding to the background information, which could be lost during transmission, and the necessity of their update. The main priority during transmission is related to the blocks representing moving objects.

All informations from blocks are coding in the XML file format chosen due to a simple transfer by Internet Trivial File Transfer Protocol (TFTP) protocol

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_37

which is a good solution for the transport of text files in the Internet, in some cases, better than HTTP for hypertext files [1].

TFTP is very simple in design and has limited features as compared to File Transfer Protocol (FTP). The User Datagram Protocol (UDP) is used by TFTP to transport data. TFTP provides no authentication and security while transferring files but is very fast end easy for implementation. Small packages of data (XML files) transmitted over the IP network allow you to adjust the flow of data to the local network load. Restrictions on transfer of technology (e.g. bitrate) of data can be included in the proposed solution.

2 Algorithm

The server side algorithm contains seven steps:

1. initialization part (also included steps 2,3,5,6,7 of algorithm),
2. frame dividing into square blocks,
3. MMC image processing for estimation of energy, histogram for each block,
4. motion detection,
5. calculation of buffer priority and compression ratio,
6. creating the XML file structure,
7. frame transmission by TFTP protocol with the defined queue priority.

The algorithm steps from the second to the seventh are repeated.

The client side algorithm:

1. Initialization and first frame collection,
2. TFTP collection data of the transmitted frame,
3. reconstruction image based on XML files and initial frame,
4. image tonal reconstruction based on stored histograms,
5. frame presentation.

The algorithm steps from the second to the fifth are repeated.

2.1 Initialization of the Server Side

During the initialization phase assuming the stability of the camera's observation parameters all the information corresponding to the whole scene is transmitted. Then, in normal working mode, only the information related to moving objects is transferred. Additionally, such data can be compressed using lossy JPEG algorithm. The choice of the TFTP protocol is motivated mainly by unproblematic realization of a buffer for image blocks and its simple implementation. Transmitted files are related to fragments of the image so for dynamic changes of image content there is a possibility of controlling the amount of data sent.

2.2 Fast Monte Carlo Image Processing

The idea of downgrading the image resolution is based on the Monte Carlo method. The image is converted into vector of length equal to the full size of image (e.g. $4 \times 4 = 16$). The new vector is prepared from randomly selected n pixels (where n is the number of draws) from the previous vector containing full information about the image. Fig. 1 presents the vector obtained for $n = 4$ draws. This vector is equivalent to the Monte Carlo image representation corresponding to the full image.

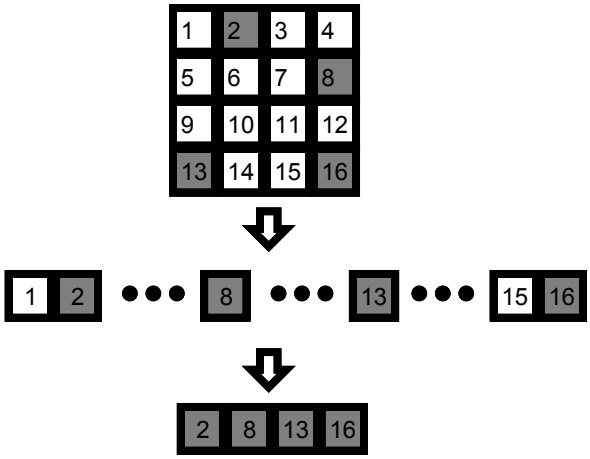


Fig. 1. Block diagram of the fast Monte Carlo image processing

The fast Monte Carlo image processing is recommended to:

- estimation of image histogram [2],
- estimation of image entropy and energy,
- estimation of threshold [3],
- binarization [4],
- objects' area estimation.

2.3 Motion Detector

Analyzed image can be divided into $T \times S$ squares of $r \times r$ pixels each. For each block the object's area using fast Monte Carlo method can be estimated independently and obtained values can be stored in an array of $T \times S$ elements. Comparing the values stored in the array for two succeeding frames it is possible to determine the changes in blocks.

Fig. 2. Server and client side video frame presentation

2.4 Transmission Queue Priority

The main priority during transmission is related to the blocks representing moving objects. The buffer is organized in accordance with the calculated energy levels for each of the blocks. The elements having the highest energy level leave the buffer as the first ones.

2.5 Data Compression Ratio

Data corresponding to blocks with detected motion are transmitted using lossy compression with adaptive change of compression ratio. The JPEG compression standard has been used and the varying compression ratio has been set as equal to the estimated (using the Monte Carlo method) value of the image block energy. For the high value of estimated energy the compression ratio is low.

2.6 Data Structure

The data structure stored in XML file contains:

- designated priority,
- signature of the video frame,
- block position in the video frame,
- the Monte Carlo histogram,
- the Monte Carlo energy,
- compressed image block.

2.7 Transmission

For one video frame each block motion detection using the fast Monte Carlo method is performed. The blocks without detected motion are marked as static. The principle of proposed method is transmission, with defined priority, only the

necessary information related to moving objects. Data from static blocks are not transmitted when the motion has been detected. In the absence of movement the static blocks are transmitted. In other cases, the image is reconstructed based on the previously transmitted image block. All data stored in the XML files are transmitted with the designated priority.

3 Simulation

The Video Server based on Windows XP has been prepared to stream data into network and Linux Debian router has been used for controlling and measuring network with unlimited bandwidth [5]. The Client has been realized on Windows XP system whereas all tested applications have been developed in Java. All systems have been implemented in the Oracle VM Virtual Box virtualization software [6]. All real and simulation tests have been performed on the previously recorded video material.

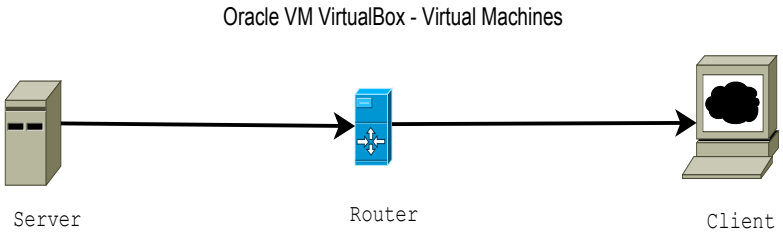


Fig. 3. Diagram of the simulation environment

The most essential aspects in the proposed algorithm is the proper choice of the block size. Using small blocks makes computations more time consuming and too large block size can easily lead to the situation when none of them is marked as static so data from previous frame cannot be utilized in processing unit. Simulation tests have been performed using four different block sizes: 16×16, 32×32, 64×64 and 128×128 pixels. Achieved results for our approach are better than those obtained using uncompressed methods regardless of the block size except 16×16 pixels. The choice of 64×64 pixels block size is the optimal solution.

4 Verification

The Video Server based on Windows 7 is prepared to stream data into network. The Client is realized on Windows 8 system. All tested applications are the same as used in simulation. The bandwidth limitation is prepared by the GPRS router on 115 kbps (technically limited by the GPRS modem). The limitations of bandwidth by ADSL (Asymmetric Digital Subscriber Line) router and Internet are not considered in this study.

Table 1. The results of simulation (data sizes in bytes for each transferred frame)

motion	uncompressed	16 × 16, n=8	32 × 32, n=16	64 × 64, n=32	128 × 128, n=64
no	304182	313322	90221	54022	82132
yes	188321	193445	23123	8944	22134
yes	161387	174553	20112	7894	20110
yes	166423	176733	21111	7965	21344

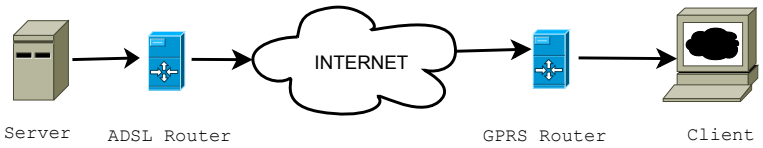


Fig. 4. Diagram of the real experimental environment

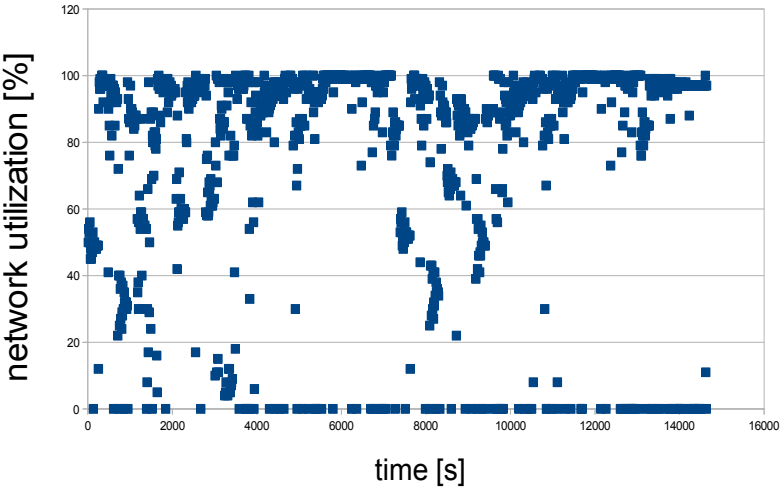


Fig. 5. Network utilization for the real transmission

The actual transmission was conducted with errors which resulted from the nature of the GSM (Global System for Mobile Communications) link as data transfer via GSM network is characterized by failures. Fig. 5 shows the states of emergency for which the link utilization is zero percent.

5 Conclusions

Presented algorithm is an efficient method of low bitrate transmission with motion detection possible to apply in the following cases: low-performance hardware

systems with limited amount of operating memory [8], changing the detection algorithm in existing systems and utilizing unused resources for estimation of additional motion parameters, real-time systems, threshold detection and histogram estimation. Possible practical applications of the proposed method are related to the following applications: security systems, Internet cameras, traffic inspection and control systems. The experiments conducted in real environment have showed that a low error rate for the GSM link is required for proper data transfer.

References

1. Ding, H., Zhang, B., Ding, Y., Tao, B.: On a novel low-cost webbased power sensor via the Internet. Sensors and Actuators A136, 456–466 (2006)
2. Lech, P., Okarma, K., Tecław, M.: A fast histogram estimation based on the Monte Carlo method for image binarization. In: Choras, R.S. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 73–80. Springer, Heidelberg (2014)
3. Oszutowska–Mazurek, D., Mazurek, P., Sycz, K., Waker–Wójciuk, G.z.: Adaptive windowed threshold for box counting algorithm in cytoscreening applications. In: Choras, R.S. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 3–12. Springer, Heidelberg (2014)
4. Lech, P., Okarma, K.: Optimization of the fast image binarization method based on the Monte Carlo approach. Elektronika Ir Elektrotechnika 20(4), 63–66 (2014)
5. Arlos, P., Fiedler, M., Nilsson, A.A.: A distributed passive measurement infrastructure. In: Dovrolis, C. (ed.) PAM 2005. LNCS, vol. 3431, pp. 215–227. Springer, Heidelberg (2005)
6. Lech, P.: A sensor network gateway for IP network and SQL database. In: Choras, R.S. (ed.) Image Processing and Communications Challenges 5. AISC, vol. 233, pp. 321–328. Springer, Heidelberg (2014)
7. Ohm, J.: Comparison of the coding efficiency of video coding standards including high efficiency video coding (HEVC). IEEE Transactions Circuits and Systems for Video Technology 22(12), 1669–1684 (2012)
8. Chen, M., Gonzalez, S., Cao, H., Zhang, Y., Vuong, S.: Enabling low bit-rate and reliable video surveillance over practical wireless sensor networks. The Journal of Supercomputing 65(1), 287–300 (2010)

Dynamic Routing and Spectrum Assignment for Varying Traffic in Flexible Optical Networks

Ireneusz Olszewski

Faculty of Telecommunications, Computer Science and Electrical Engineering,
University of Technology and Life Sciences,
Al. Prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland
Ireneusz.Olszewski@utp.edu.pl

Abstract. The problem of service time-varying traffic in flexible transparent optical networks, assuming that the incoming connections and their duration are not known in advance, is considered in this work. Time-varying traffic requires dynamic spectrum allocation for connections implemented in the network. In the considered problem, a path with the required spectrum around the reference frequency is determined for each incoming connection by the routing algorithm. In order to take into account the dynamic spectrum allocation for the connections in the network, two spectrum expansion/contraction schemes have been applied on the basis of which the average blocking probability for the incoming request additional slot was determined.

1 Introduction

A spectrum-sliced elastic optical path network (SLICE) has been proposed as an efficient solution for flexible bandwidth allocation in optical networks. The applied optical orthogonal frequency -division multiplexing (OFDM) supports the transport of the multi-granularity Internet traffic, however, it requires a grid-flexible (sliced or mini-grid) or fully gridles network. In SLICE, the spectra of OFDM signals are flexible and sliced into any continuous spectrum slots transferring traffic with any speed transmission. Similarly to Routing and Wavelength Assignment Problem in DWDM networks, Routing and Spectrum Assignment Problem (RSA) also appears in the SLICE networks. In [1] the dynamic problem of RSA, which takes into account the relationship between the spectrum bandwidth, modulation format and traffic bitrates, has been formulated. The objective function includes minimizing the path length, and spectrum continuity constraints and non-overlapping spectrum assignment constraints which play the role of constraints in the formulated optimisation problem. To solve this problem, two different algorithms based on the segment representation of the spectrum, which naturally support both mini-grid and fully gridless networks, have been proposed. Whereas, in [2] a dynamic routing algorithm with adaptive modulation in distance SLICE networks has been proposed. In [3] the RSA problem has been formulated as an integer linear programming task and a heuristic algorithm for solving this problem has been proposed, if the ILP solution is not

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_38

achievable. In [4] a k-path Signaling-based RSA scheme has been proposed and simulation results show that in Flexible Bandwidth Optical Networks it performs better than other RSA schemes. In turn, in [5] the problem of planning an optical network based on OFDM, where connections are protected by the spectrum non-overlap rule, has been formulated. To solve this problem there have been proposed several algorithms including optimal and decomposition ILP algorithms and a sequential heuristic algorithm combined with appropriate ordering policies and simulated annealing meta-heuristic.

In [6] the problem of serving time-varying traffic in a spectrum flexible optical network is considered, where the spectrum allocated on the connections changes dynamically with time so as to follow the required source transmission rate. For a description of these spectrum changes, two spectrum expansion/contraction schemes have been proposed, basing on which, a model determining the probability of the network has been developed. It should be emphasized that in the considered problem the number of connections with the number of slots required for each of them is known in advance and the connections are only set up (long-lived connections).

In this paper, the RSA problem has been formulated, in which a stream of request connection is random with random duration of these connections (short-lived connections) and the number of slots in the connections varies dynamically in order to follow the source transmission rate. Furthermore, an algorithm solving this problem has been proposed.

It should be noted that further searching of algorithms solving the dynamic RSA problem, where connections come randomly with random duration, is necessary.

The remaining part of this paper is as follows: the first part contains a formulation of the optimization problem, the second part contains the proposed algorithm solving the formulated RSA problem, assuming that connections with the required number of slots are not known in advance. The third part presents the obtained results, and the fourth one contains a summary and conclusions.

2 Formulation of the Considered Problem

Let $G(N, L)$ be the network, where N is a set of nodes, and L is a set of unidirectional lines l, such that $l \in L$. The spectrum of each link l is divided into T slots (numbered from 0 to $T-1$) with a granularity equal to 12.5 or 6.25 GHz. R is the symbol rate (in baud), and G (Hz) is a guard band between adjacent connections. Furthermore, let the current request be for C units of bandwidth [in b/s] between a pair of nodes: $s, d \in N$. The relationship between the bit rate of signal C and the spectrum of signal B in case of OFDM modulation can be described as follows [1]:

$$B = (\lceil C/2mR \rceil + 1)\, R \tag{1}$$

where m is the number of bits per one symbol, and R is the symbol rate (in baud) for each of the subcarriers. The above relationship was obtained assuming that all the subcarriers have the same signal format: PDM and the AMF. The

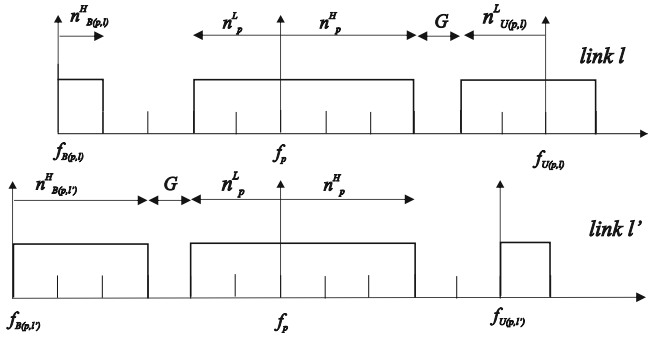


Fig. 1. Using the spectrum of slots by connection p and the adjacent connections on links l and l' at the same time

required path p between the pair of nodes (i, j) with the required number of slots $n_p = n_p^L + n_p^H$ around the reference frequency f_p for random incoming request of the connection is determined by the RSA algorithm. Duration of the connection is also random. It should be noted that in order to satisfy the spectrum continuity constraint, the same slots must be assigned for connection p on all links l, $l \in p$. Moreover, for dynamic RSA problem, the required number of slots of the connection varies in time so as to follow changes source transmission rate while maintaining non-overlapping spectrum assignment constraint, which means that two or more connections can not use the same slots on any link at the same time. Fig. 1 [6] shows the spectrum slots utilization on two links l and l', belonging to the path p. Let $B(p, l)$ and $U(p, l)$ denotes the bottom and upper spectrum-adjacent connectins on link $l \in p$. Whereas, let $f_{B(p,l)}$ and $n^H_{B(p,l)}$ be the reference frequency and the number of upper slots of the spectrum (above the reference frequency) for bottom adjacent connection to connection p on link l. In turn, let $f_{U(p,l)}$ and $n^H_{B(p,l)}$ be the reference frequency, and the number of lower slots of the spectrum (below the reference frequency) for upper adjacent connection to connection p on link l. G is a guard band between the adjacent connections on link l. In case of traffic fluctuations in time, the volume of spectrum for connection p can be increased by additional slots reservation or reduced by releasing some slots. The range of variability of the number of slots n_p^L and n_p^H around the reference frequency f_p, while maintaining the non-overlapping spectrum assignment constraint, can be determined as follows [6]:

$$0 \leq n_p^L \leq f_p - \max_{l \in p} \left(f_{B(p,l)} + n^H_{B(p,l)} \right) - G \tag{2}$$

$$0 \leq n_p^H \leq \min_{l \in p} \left(f_{U(p,l)} - n^L_{U(p,l)} \right) - f_p - G \tag{3}$$

To determine the impact of the traffic dynamic changes on the network blocking probability in [6], there have been proposed two spectrum allocation (Expansion/Contraction) schemes, on link l belonging to connection p. In the first

scheme, named as *Constant spectrum allocation (CSA) scheme*, the slots for connection p can be used from reference frequency f_p to reference frequency $f_{U(p,l)}$ of adjacent connection p on link l. Therefore, the range of slots variation n_p^H used in this scheme can be described as:

$$0 \leq n_p^H \leq N_p^H \tag{4}$$

$$\text{where } N_p^H = \min_{l \in p} \left(f_{U(p,l)} \right) - f_p - G \tag{4a}$$

Thus, in this scheme, no slots can be shared between the adjacent connections on the links of the network. If the request of additional slot appears for connection p and the number of slots n_p^H in this time is equal to N_p^H, then this request will be blocked. Thus, this scheme can only be used for comparative purposes. In the second scheme, named as *Dynamic High Expansion-Low Contraction (DHL) scheme*, slots for connection p can be used both below and above the reference frequency f_p. If the transmission rate for connection p increases, additional slots on link l are occupied, first above frequency f_p increasing n_p^H up to the slots used in a given time by the upper adjacent connection, i.e.:

$$0 \leq n_p^H \leq N_p^H \tag{5}$$

$$\text{where } N_p^H = \min_{l \in p} \left(f_{U(p,l)} - n_{U(p,l)}^L \right) - f_p - G \tag{5a}$$

In the absence of free slots, the requested slots for connection p are occupied below frequency f_p increasing $\frac{L}{p}$ up to the slots used in a given time by the bottom adjacent connection, i.e.:

$$0 \leq n_p^L \leq N_p^L \tag{6}$$

$$\text{where } N_p^L = f_p - \max_{l \in p} \left(f_{B(p,l)} - n_{B(p,l)}^H \right) - G \tag{6a}$$

In the absence of free slots the request will be blocked.

After determining the schemes of using the free slots, the considered problem in this paper can be defined as the problem of RSA, in which the stream of arrival connections is random with random duration of these connections and the required number of slots in the connections is dynamic - varying in time, so as to follow the required source transmission rate.

The solution to this problem involves determining an algorithm that for every request incoming to the network would determine the path and the reference frequency, together with the required number of slots, taking into account the possibility of slots sharing, so as to minimize the average blocking probability.

In [6] to find a solution for the dynamic (time-varying) RSA problem, in which the set of connections with a designated number of slots for each of them is known in advance, the heuristic algorithm is proposed. Assuming that the transmission

system supports T slots on each link (fiber) and knowing the volume of traffic for each connection, the dynamic (time-varying) RSA problem is transformed into a static RSA problem. Then, the static RSA problem is solved by a heuristic algorithm based on Simulated Annealing [7]. After solving the static RSA problem (resulting in the path p is obtained with the reference frequency fp for each connections), the blocking probability is determined using the presented schemes spectrum expansion/contraction in case of changes (increase/decrease) in the source transmission rate on the connections. Then, for the same set of connections and a number of slots for each of them, next iteration is performed. The algorithm terminates after K acceptable solutions. A solution that provides the minimum blocking probability of the network from the set of K solutions is the final solution.

It should be emphasized that the designation of the solution that minimizes the blocking probability is based on repetition of the algorithm for the known in advance set of connections with the required number slots for each of them.

In this paper the analysis of the network using both spectrum allocation schemes and taking into account the intensity of the arrival of requests for additional slots in the existing connections has been carried out. The general idea of the analysis of the network is as follows: the connection requests with C units of bandwidth (in bit/s) uniformly distributed from $C_{\min}$ to $C_{\max}$ arrive to the network consisting of n nodes. A stream of requests between each pair of nodes is the Poissonian, with intensity λ (for simplicity the same for each pair of nodes), and the connection duration has an exponential distribution with the mean equal to $1/\mu = 1$. Therefore, the average traffic (in bps) between each pair of nodes is equal to $\rho\bar{c}$, where $\rho = \lambda/\mu$ is the traffic in Erl. Implementation of C units of bandwidth for the incoming request requires the signal spectrum, defined by formula (1). After taking into account the grid with granularity equal to 6.25 or 12.5 GHz, the required number of slots for a given connection can be determined. Then, to solve the formulated RSA problem, algorithm 2 in [8], which is a modification of the MSP algorithm (Modified Shortest Path Algorithm) [1] is used. The algorithm in [1] has been proposed to solve the static RSA problem where the stream of requests is Poissonian and the connection duration has an exponential distribution. In [1,8], a spectrum segmented representation which fully supports the coarse mesh WDM network, a mini-grid networks (slot-based) and ideal fully gridless networks, have been used. The algorithm used in this paper designates path p and reference frequency f_p together with the required number of slots for each incoming request between the pair of nodes (i, j). It should be noted that the designated slots must satisfy the spectrum continuity constraint and the non-overlapping spectrum assignment constraint. Moreover, guard band G (in slots) is allocated between adjacent connections to eliminate interference. After allocating a specified number of connections (where each of them requires the number of slots based on (1)), the network given state is obtained. The set of feasible states of the network is determined by the occupied slots on the links by the connections and is described by (2) and (3). For a given state of the network, an analysis of the network is made according to the assumption that

the required number of slots for the connections dynamically vary in time so as to follow the source transmission rate.

3 The Solution of the Considered Problems Taking into Account the Slots Allocation Schemes

The network dynamic operation is achieved by associating Poissonian stream of requests of additional slots with intensity λ_p with each connection p, in a given state of the network. The duration of the slots is exponentially distributed with the mean equal to $1/\mu_p = 1$. In order to simplify, it was assumed that the arrival intensity of the additional slots for all the connections in a given state of the network is the same and equal to λ_p. The defined problem is solved for both shown slots allocation schemes.

3.1 Solving the Problem for Constant Spectrum Allocation (CSA) Scheme

In the constant spectrum allocation scheme the slots are occupied for connection p above the reference frequency f_p of the interval defined by (4). Thus, the incoming request allocation of additional slots for this connection encounters $N_p^H - n_p^H$ of free slots. Rejection of this request will appear, when $n_p^H = N_p^H$. Thus, the blocking probability for the incoming request on the path p can be determined as first Erlang function: $b_p = E_{1,N_p^H - n_p^H}(a_p)$, where $a_p = \lambda_p/\mu_p$. Knowing the connections in a given state of the network and the blocking probability of requests for additional slots for each of the connection it is possible to determine the average probability in the network.

3.2 Solving the Problem of Dynamic High Expansion-Low Contraction (DHL) Scheme

On the other hand, in case of DHL schema, the required slots for connection p are occupied both above and below the reference frequency f_p. Above frequency f_p, additional slots are occupied from the range $\left(n_p^H - N_p^H\right)$, where N_p^H is defined by equation (5a). In order to simplify the analytical model it was assumed that all connections in the network, except the connection on path p, occupy the slots above the reference frequency. Thus, the blocking probability for the incoming request additional slot from $N_p^H - n_p^H$ free slots above the frequency f_p, where $N_p^H = \min_{l \in p}\left(f_{U(p,l)}\right) - f_p - G$ can be described as $b_p^H = E_{1,N_p^H - n_p^H}(a_p)$, where $a_p = \lambda_p/\mu_p$. If the request is blocked, it will be directed to the free slots below frequency f_p. In this case, however, according to the assumption adopted above, for all links of connections p, $l \in p$, the slots lying below f_p will be affected by slots of the bottom adjacent connections. Therefore, each link $l \in p$ should be considered separately, which is a further simplification. The blocking probability request of the free slot for connection p on link $l \in p$ below f_p can be defined as

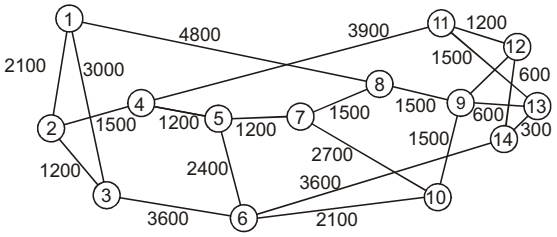


Fig. 2. Topological structure of the network, NSFNET [1]

$b^L_{p,l} = E_{1,N^L_{p,l}-n^L_{p,l}}\left(a_p(1+b^H_p)\right)$, where $N^L_{p,l} = f_p - \left(f_{B(p,l)} + n^H_{B(p,l)}\right) - G$. Thus, the average probability of the blocking request additional slots on the path p can be determined as $b_p = b^H_p\left[1 - \prod_{l\in p}\left(1 - b^L_{p,l}\right)\right]$. Knowing the connections for some state of the network and the probability of blocking request for additional slot for each of them the average probability of blocking request for additional slots in the network can be determined.

4 Obtained Results

Verification of the considered models was performed for the network, whose topological structure is shown in Fig. 2 [1]. It consists of 14 nodes connected by links and each of them includes T slots. Each edge in this graph is a pair of unidirectional links. In this work it is assumed that each node may be an input and output node. Thus, 196 pairs of nodes in the network can be distinguished. The lengths of the links are shown on the edges of the graph. The network simulation was carried out basing on Monte Carlo method. It was assumed that the stream of connection requests between each pair of nodes (s, d) is Poissonian with intensity $\lambda = 2$. The duration of the connections is exponentially distributed with mean $1/\mu = 1$. Bandwidth of the connections is uniformly distributed from 50 Gbps to 200 Gbps with mean equal to = 125 Gbps. The intensity of the requests for additional slots for each connection p is a parameter and the duration of these slots is exponentially distributed with mean $1/mu_p = 1$. Simulation of the network was performed under dynamic conditions, it means that the connections are set up and released (short-lived connections). Then, for the state of the network, which also depends on the number of the connections, a simulation of the dynamic changes of the spectrum (in slots) is performed for all connections. Registration of the rejected requests additional slots is performed after obtaining the system steady state. It should be noted that both the analytical and simulation models take into account the adopted spectrum allocation schemes and these models are carried out for the same state of the network. For the input data, the obtained results are averaged on the basis of 30 simulation (30 states of the network for the same data input). Furthermore, it was assumed that OFDM

Table 1. The blocking probability depending on the intensity of arriving requests

$\sum \lambda_p$	$\lambda = 2, \mu = 1, m = 2,$ $\bar{C}$ = 125 Gbps, R=1 Gbaud, G=1, T=300			
	CSA		DHL	
	Analytical model	Simulation	Analytical model	Simulation
3.8E+01	0.5379±0.02238	0.3568±0.01997	0.2630±0.01560	0.3372±0.02540
7.5E+01	0.5404±0.01304	0.3811±0.01589	0.2684±0.01660	0.3356±0.01495
1.5E+02	0.5630±0.01354	0.3918±0.01795	0.2860±0.01448	0.3446±0.02107
2.2E+02	0.5824±0.01261	0.4187±0.01759	0.3177±0.01918	0.3779±0.02159
2.9E+02	0.5878±0.01760	0.4262±0.02566	0.3391±0.02006	0.3895±0.01474
3.6E+02	0.5958±0.01075	0.4644±0.01903	0.3743±0.01318	0.4204±0.01800
4.6E+02	0.6080±0.02063	0.4717±0.02230	0.3765±0.02386	0.4233±0.02469
4.9E+02	0.6194±0.01372	0.4803±0.01826	0.3887±0.01652	0.4280±0.01705
5.2E+02	0.6299±0.01146	0.4985±0.01353	0.4157±0.02158	0.4604±0.01967
6.3E+02	0.6490±0.01297	0.5235±0.02429	0.4312±0.01714	0.4752±0.01972
7.3E+02	0.6564±0.00920	0.5335±0.01288	0.4407±0.01507	0.4789±0.01728

Table 2. The blocking probability depending on the number of slots provided by a transmission system on each link

T	$\Lambda = 2, \mu = 1, m = 2,$ $\bar{C}$ = 125 Gbps, R=1 Gbaud, G=1, $\sum \lambda_p$=3.6E+02			
	CSA		DHL	
	Analytical model	Simulation	Analytical model	Simulation
300	0.5958±0.01075	0.4644±0.01903	0.3743±0.01318	0.4204±0.01800
320	0.6007±0.02428	0.4420±0.02336	0.3555±0.02736	0.4190±0.02586
340	0.5887±0.01504	0.4495±0.01026	0.3571±0.00991	0.4180±0.01928
360	0.6077±0.01431	0.4580±0.01948	0.3732±0.01443	0.4214±0.01699
380	0.6058±0.02169	0.4564±0.03022	0.3730±0.02544	0.4270±0.02927
400	0.5896±0.01417	0.4422±0.01931	0.3614±0.01481	0.4249±0.01158

systems have the same symbol rate equal to $R = 1$ Gbaud, and the guard band is equal to $G = 1$ slot. Table 1 shows the average blocking probability of additional slot request depending on the intensity of the arrival of additional slots requests for the connections in the network with a fixed number of slots $T = 300$ realized by a transmission system on each link. The obtained results prove that the probability of blocking additional slot request using the CSA scheme is the upper bound for the results obtained using DHL scheme for the whole considered range of λ_p. This results from the fact that in case of DHL scheme slots are shared by the adjacent connection as opposed to the CSA. Moreover, the blocking probability of additional slot request is surprisingly large compared to the results in [6]. It should be emphasized that these two RSA problems are completely different. In the considered dynamic RSA problem in [6], the set of connections with the required number of slots is known in advance, while in the dynamic RSA problem considered in this work, the stream of incoming requests to the network is Poissonian and the connections are not known in advance.

In turn, Table 2 shows the blocking probability of additional slots request depending on the number of slots implemented by the transmission system on each link of the network. As can be seen, even with a large increase in the number of slots (over 30%), implemented on the links of the network, the blocking probability of requests is practically unchanged. This is caused by the applied algorithm proposed in [1] and revised in [8] for solving the static RSA problem, in which a stream of requests is Poissonian and duration of the connections is exponentially distributed (the number o connections is not known in advance). The algorithm is based on the classical Dijkstra algorithm [9], which is an avid algorithm, which means that in every step it makes a selection of locally optimal decision, regardless of the situation in the next step. Therefore, to solve the defined problem in this work we should go in the direction of heuristic algorithm with prediction, which during determining the path p for incoming request could predict the ability to dynamically change the number of the required slots (taking into account the sharing of slots) minimizing the blocking probability of additional slots.

The thesis accepted at the beginning of this paper confirms that it is necessary to search for algorithms for solving dynamic RSA problem, in which a stream of connections and service time are not known in advance. Furthermore, the connections implemented for the assumed input data require from 3 to 9 slots. Thus, the realization of the request for additional slot increases the amount of traffic (in slots) from 11% to 33% for the connection. Therefore, for a given state of the network, in which the specified number of connections chosen by avid algorithm are realized the request of additional slots is rejected with quite a high probability.

5 Conclusions

The paper presents a dynamic time-variable RSA problem, in which a stream of requests and duration of the connections are random (connections are not known in advance). To solve this problem, two schemes of slots allocation on the connections have been used in order to follow the required source transmission rate. The first one does not take into account the possibility of sharing slots between adjacent connections on links of the network and it is presented for comparative purposes only. The second scheme enables sharing slots between adjacent connections taking into account the spectrum continuity constraints and the spectrum non-overlap constraints. To solve the formulated problem in this paper we transform this problem into a static problem of RSA, in which a stream of requests and connections duration are random. Then, after obtaining a given state of the network, requests for additional slots for connections are generated taking into account slots allocation schemes. To solve the static RSA problem the algorithm based on Dijkstra algorithm, which is an avid algorithm, has been used. Moreover, to make a comparison of analytical and simulation models (for each considered slot allocation scheme), these models were used for the same network state. The obtained results prove that the blocking probability

of additional slots requests obtained on the basis of DHL scheme, for both the analytical and the simulation model, are upper bounded by results obtained on the basis of the CSA schema. This shows that in case of DHL scheme, there is spectrum (in slots) sharing for both the analytical and the simulation model. Large values of the blocking probability result from the algorithm used to solve the static RSA problem. Simultaneously, these results indicate the direction of further studies to determine the heuristic algorithms with prediction, solving the dynamic RSA problem, in which the incoming request and the call duration are random (not known in advance).

References

1. Wan, X., Hua, N., Zheng, X.: Dynamic Routing and Spectrum Assignment in Spectrum -Flexible Transparent Optical Networks. Opt. Commun. Netw. 4(8), 603–613 (2012)
2. Jnno, M., Kozicki, B., Takara, H., Watanabe, A., Sone, Y., Tanaka, T., Hirano, A.: Distance adaptive spectrum resource allocation in spectrum-sliced elastic optical path network. IEEE Communication Magazine 48(8), 138–145 (2010)
3. Klinkowski, M., Walkowiak, K.: Routing and Spectrum Assignment in Spectrum Sliced Optical Path Network. IEEE Communications Letters 15(8), 884–886 (2011)
4. Jin, Q., Wang, L., Wan, X., Zheng, X., Zhou, B., Liu, Z.: Study of Dynamic Routing and Spectrum Assignment Schems in Bandwidth Flexible Optical Networks. In: Communicationa and Photonics Conference, Asia (2007, 2011)
5. Christodoulopoulos, K., Tomkos, I., Varvarigos, A.E.: Routing and Spectrum Allocation in OFDM-based Optical Networks with Elastic Bandwidth Allocation. In: Telecommunications Conference 2010, December 6-10. IEEE (2010)
6. Christodoulopoulos, K., Varvarigos, E.: Routing and Spectrum Allocation Policies for Time Varying Traffic in Flexible Optical Networks. In: 16th International Conference on Optical Network Design and Modeling (ONDM), April 17-20 (2012)
7. Christodoulopoulos, K., Tomkos, I., Varvarigos, A.E.: Elastic Bandwidth Allocation in Flexible OFDM-based Optical Networks. J. Lightwave Technology 29(9) (2011)
8. Olszewski, I.: Algoritms of Routing and Spectrum Assignment in Spectrum Flexible Transparent Optical Networks. Przegląd Elektrotechniczny 89, 248–253 (2013) ISSN 0033-2097
9. Sysło, M.M., Deo, N., Kowalik, J.: Discrete Optimization Algorithms: With Pascal Programs. Dover Publications, Mineola (2006)

Reduction of Reference Set for Network Data Analyzing Using the Bubble Algorithm

Artur Sierszeń

Institute of Applied Computer Science, Lodz University of Technology
artur.sierszen@p.lodz.pl

Abstract. Data in modern networks should be analysed in real time. This is necessary to maintain security and enable management. The real problem is the size of analysed data and its classification. One of several solutions to these problems may be reduction the reference network data set. Most algorithms for the condensation of the reference set involve a lot of computation when processing a very large set which contains several dozens of objects. That was the basis for our attempt to develop a completely new classifier which would maintain the quality of classification on the levels obtained with the primary reference set as well as allow to accelerate computations considerably. The proposed solution consists in covering the primary reference set with disjoint hyperspheres; however, these hyperspheres may contain objects from one class only. Classification is completed when it has been determined that the classified point belongs to one of the mentioned spheres. If an object does not belong to any hypersphere, it is counted among the objects of the same class, to which the objects from the nearest hypersphere belong (the distance to the centre of the sphere minus the radius). As was shown in our tests, this algorithm was very effective with very large data sets.

1 Introduction

What is most wanted in a classifier is a low share of incorrect decisions (high quality of classification). Speed of classification is also very important in some applications, e.g. when a classifier analyses optic images.

When classifying very large sets, classifiers using distance functions should be taken into account, among many other solutions (including neural networks or decision trees). This group of algorithms contains classifiers with not too strict assumptions. They provide the quality of classification similar to that of the Bayes classifier, which is theoretically the best [1]. This type of classifiers includes e.g. the *k nearest neighbours* rule (*k-NN*) [2] and its many modifications, e.g.:

- the concept of the surrounding neighbourhood and the resulting *k-NCN* decision rule (*k - Nearest Centroid Neighbours*) [3],
- the *k-DNN* rule (*k-Diplomatic Nearest Neighbours*) [4],
- local metrics [5,6],

© Springer International Publishing Switzerland 2015

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_39

- the concept of the *k-ENN* rule (*k-Edited Nearest Neighbour*) [7],
- the *k-NN* rule with a pre-classifier of *1-NN* type [8],
- the fuzzy *k-NN* rule [9,10].

The *k-NN* rule assigns the classified object to the class which is represented by most of its k nearest neighbours in the reference set. The reference set is usually the whole learning set.

2 Speed of *1-NN* Classification Algorithm

In practical applications, a significant role is played by the classifier operating based on the nearest neighbour rule, which is a particular case of the *k-NN* rule. It is the fastest type of the *k-NN* rule and, what is more, a classifier of this type may be used to approximate other versions of the *k-NN* rule. To this purpose, the reference set must be re-classified using an approximated classifier. Then, the *1-NN* rule should be applied with the modified reference set. So far, many algorithms for the reduction of the reference set have been proposed in order to accelerate the *1-NN* rule. New algorithms are being developed all the time. Another way to accelerate the nearest neighbour rule is to condense the reference set, i.e. to create a set of artificial objects based on the primary learning set. Only few methods of reference set condensation have been developed. Compared to the algorithms for the reduction of the reference set, they are able to control the compromise between the size of the obtained condensed set, i.e. the speed of classification, and the quality of classification, measured with the percentage of correct decisions.

3 Bubble Algorithm

The presented algorithm for the condensation (reduction) of the reference set involves the modification of the decision rule. Therefore, the condensation process itself may be considered as the learning stage. Since it is necessary to change the classification rule, the classification phase requires detailed description as well as the learning phase, which consists in the condensation of the reference set. The classification phase does not use the standard version of the *1-NN* rule.

In the learning phase, the separated part of the set of objects with the known class affiliation, i.e. the learning set, is used to determine a set of hyperspheres. Each of the hyperspheres contains objects from the learning set which belong to one class only. Such hyperspheres are described as homogenous.

An homogenous hypersphere is described with the following parameters:

- the base point a point from the learning set which is the centre of an homogenous hypersphere;
- the radius a distance from the base point to the nearest point belonging to another class or to the edge of another homogenous area;

– the number of contained points which determines how many points belonging to the same class as the base point (excluding the base point) is contained in an homogenous area.

The process of constructing homogenous hyperspheres is presented in the Fig. 1 and Fig. 2. Its graphic interpretation gave the author an idea of the name for the developed method.

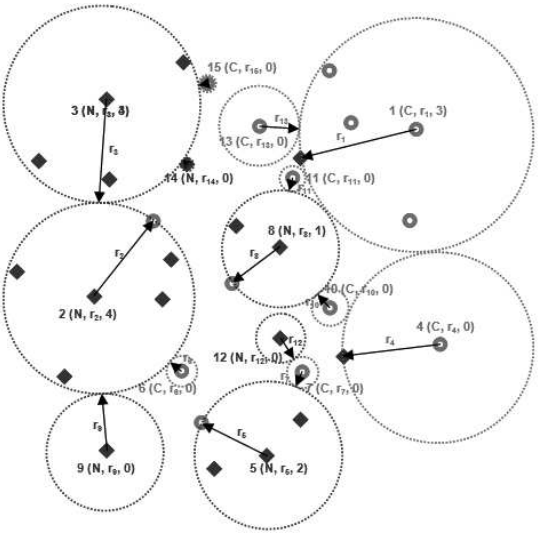


Fig. 1. The learning phase of the bubble classifier the construction of homogenous hyperspheres (a construction diagram)

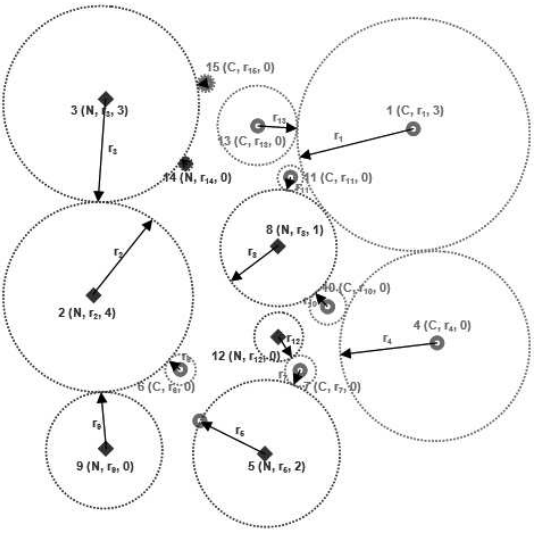


Fig. 2. The learning phase of the bubble classifier the construction of homogenous hyperspheres (after the internal objects have been removed from hyperspheres)

The algorithm used in the learning phase of the bubble classifier chooses a random point from the learning subset (or, at the beginning, from the whole learning set) and marks it as the base point. Then, the distance is determined from the base point to the nearest point belonging to another class or to the edge of another homogenous area. This distance is used to determine the range of a homogenous hypersphere. Points lying inside that area are sought, their number is remembered, and then these points are deleted. The point from another class which was used to determine the range of the area is deleted as well. The learning set is covered with hyperspheres. This algorithm is presented below in the form of the pseudocode.

START; i = 2; $T^1 = T$;
01. Choose a random point t_1 from T^1 set, the distance r_1 to the nearest point y_1 from another class and the number n_1 of points from T set located in the hypersphere K^1=(t_1,r_1,n_1);
02. $T^i = T^{i-1}Z^{i-1}\{y_{i-1}\}$, if y_{i-1} does not exist, assume that $\{yi-1\}$ is a empty set;
03. If T^i is empty, go to 06.;
04. Choose a random point t^i from T^i set and determine K^i=(t_i, r_i, n_i):
    - distance d_1 from t_i to the nearest point y_i from another class;
    - distance d_2 to the nearest, already existing, hypersphere K^j, j=1,2,..,i (it is computed as a distance to the centre of the sphere minus its radius);
    - mark the shorter of these two distances as r_i;
    - determine a number of points n_i from T^i set located in the hypersphere with t_i centre and r_i radius;
05. i=i+1; go to 02;
06. END

where:

T - the learning set containing m objects;
T^i - sets reduced in individual iterations;
K^i=(t_i,r_i,n_i) - a combination: a centre, a radius, a number of points inside a hypersphere without the base point;
Z^i - a set of points from T set located inside Ki sphere.

As a result of the learning phase of the bubble classifier, the set of homogenous hyperspheres is created. The next step consists in sorting them. The first criterion of choice is the length of the radius which determined the area (from the longest one to the shortest one). If several areas have the same radius, the areas with a greater number of reduced points are chosen first. The sorting is meant to accelerate the next phase of the bubble classifiers operation, i.e. the classification phase.

Points are loaded from the set of objects for classification or from the testing set (in the Fig. 3, these are represented by stars). After that, their distribution in reference to the determined homogenous hyperspheres is examined. Three cases are possible:

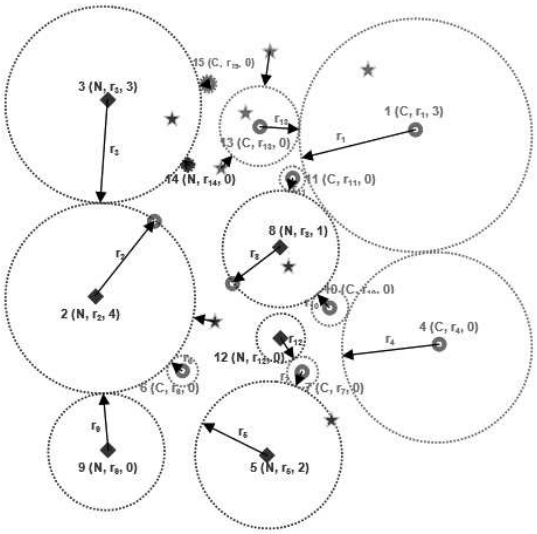


Fig. 3. The classification phase of the bubble algorithm

- a point lies inside an homogenous area the point is assigned to the class which the area is assigned to;
- a point lies on the edge of an homogenous area the point is assigned to the class which the hypersphere is assigned to;
- a point lies outside each homogenous hypersphere the point is assigned to the class which the nearest hypersphere is assigned to (this is determined based on the distance to its edge).

4 Experimental Results

The bubble classifier was implemented in C++ in Microsoft Visual Studio 2012 environment and it was tested in Windows 7 Professional Service Pack 1 (64-bit) environment with the use of a PC computer equipped with Intel(R) Core(TM) i3-2370M 2,40 GHz and 8 GB of operating memory.

4.1 Subsection Heading

During the experiments and tests of the algorithm and during the verification of its operation, the author focused mainly on the NETWORK set. This set was made based on the KDD Cup Data [11]. t is very large. A general description of the set:

- number of classes: 23,
- number of properties: 41,
- number of samples: 494021.

The classes represent the type of data sent in the network. The raw training data was about four gigabytes of compressed binary TCP dump data from seven weeks of network traffic. This was processed into about five million connection records. Similarly, the two weeks of test data yielded around two million connection records. A connection is a sequence of TCP packets starting and ending at some well defined times, between which data flows to and from a source IP address to a target IP address under some well defined protocol. Each connection is labeled as either normal, or as an attack, with exactly one specific attack type. Each connection record consists of about 100 bytes. Attacks fall into four main categories:

- DOS: denial-of-service, e.g. syn flood;
- R2L: unauthorized access from a remote machine, e.g. guessing password;
- U2R: unauthorized access to local superuser (root) privileges, e.g., various "buffer overflow" attacks;
- PROBE: surveillance and other probing, e.g., port scanning.[12].

Table 1. Size of each class in the Network set

Class number	Class name	Attack category	Number of elements
1	back	DOS	2203
2	buffer_overflow	U2R	30
3	ftp_write	R2L	8
4	guess_passwd	R2L	53
5	imap	R2L	12
6	ipsweep	PROBE	1247
7	land	DOS	21
8	loadmodule	U2R	9
9	multihop	R2L	7
10	neptune	DOS	107201
11	nmap	PROBE	231
12	normal	-*	97278
13	perl	U2R	3
14	phf	R2L	4
15	pod	DOS	264
16	portsweep	PROBE	1040
17	rootkit	U2R	10
18	satan	PROBE	1589
19	smurf	DOS	280790
20	spy	R2L	2
21	teardrop	DOS	979
22	warezclient	R2L	1020
23	warezmaster	R2L	20

* Normal network data (not from network attack).

Table 2. Operation results (the average value, median, standard deviation, and the minimum and maximum value) of the bubble classifier algorithm obtained for the Network set

The division into the learningtesting set		Measured value	Average	Median	Standard devision	Min. value	Max. value
test set[%]	10	measure[ms]	3542553	3540127	16410	3511195	3565009
learn set[%]	90	error comp.[ms]	171721	171719	18	171696	171750
test set[pcs]	49394	correct	8040	8047	22	8001	8069
learn set[pcs]	444627	incorrect	156	150	22	127	195
test set[%]	20	measure[ms]	4192530	5199793	290810	4650072	5824598
learn set[%]	80	error comp.[ms]	955381	1191273	34201	1124309	1260279
test set[pcs]	98795	correct	78880	98605	17	98566	98628
learn set[pcs]	395226	incorrect	156	191	17	167	229
test set[%]	30	measure[ms]	3030579	3808139	189459	3431507	4180074
learn set[%]	70	error comp.[ms]	1237541	1551423	40384	1475966	1639126
test set[pcs]	148198	correct	118316	147897	20	147852	147930
learn set[pcs]	345823	incorrect	242	301	20	268	346
test set[%]	40	measure[ms]	2178325	2725863	157376	2404775	3094827
learn set[%]	60	error comp.[ms]	1431747	1792709	50276	1696799	1908991
test set[pcs]	197600	correct	157732	197173	40	197050	197222
learn set[pcs]	296421	incorrect	348	427	40	378	550
test set[%]	50	measure[ms]	1565168	1961391	159910	1623230	2171321
learn set[%]	50	error comp.[ms]	1553159	1944356	79132	1775767	2047519
test set[pcs]	247016	correct	197134	246413	35	246357	246472
learn set[pcs]	247005	incorrect	470	593	35	533	648
test set[%]	60	measure.[ms]	1002840	1214190	101776	1140081	1476823
learn set[%]	40	error comp.[ms]	1527153	1888961	74686	1827059	2074866
test set[pcs]	296404	correct	236457	295588	76	295416	295680
learn set[pcs]	197617	incorrect	667	817	76	724	988
test set[%]	70	measure[ms]	547902	679772	43013	615031	751063
learn set[%]	30	error comp.[ms]	1331126	1655226	56110	1571672	1747141
test set[pcs]	345806	correct	275766	344722	110	344459	344867
learn set[pcs]	148215	incorrect	879	1084	110	939	1347
test set[%]	80	measure[ms]	241222	302235	27974	253906	356819
learn set[%]	20	error comp.[ms]	974388	1211592	64330	1114201	1353958
test set[pcs]	395209	correct	314836	393561	154	393169	393771
learn set[pcs]	98812	incorrect	1332	1648	154	1438	2040
test set[%]	90	measure[ms]	49635	57681	13243	46644	93366
learn set[%]	10	error comp.[ms]	451740	547514	59931	493491	702719
test set[pcs]	444610	correct	353252	441637	341	440930	442172
learn set[pcs]	49411	incorrect	2436	2974	341	2438	3680

In the NETWORK set, there are 494021 samples:

- 190065 TCP;
- 20354 UDP;
- 283602 ICMP.

The properties of the sample represent some information about the packet eg. type of protocol, flag, service source and destination IP etc. All properties described with a symbol have been replaced with a number.

4.2 Sequence of Set Tests

All tests were repeated 20 times. Each series was repeated for a different proportion of division into the testing set and the learning set (the following proportions were considered: 1 to 9; 2 to 8; 3 to 7; 4 to 6; 5 to 5; 6 to 4; 7 to 3; 8 to 2; 9 to 1). As a result, it was possible to thoroughly examine the influence of the size of the learning and the testing sets on the classification quality and, above all, on the speed of algorithms operation and computation of classification error. Table 2 presents the results (the average value, median, standard deviation, and the minimum and maximum value) of the bubble algorithm classifier for 20 measuring series, repeated for nine different divisions of the Network set into the learning set and the testing set.

5 Discussion

Fig. 4 presents the comparison of the average values obtained for the specific size of the learning set and the testing set. Additionally, the classification error obtained with the 1-NN method is shown, a factor which makes it easier to compare the quality of performed classification. In the figure, the x axis determines the proportion of the testing set to the learning set.

As can be seen in Fig. 4, in all test series for this set, the algorithm obtained a worse result of classification than in the case of the classification with the 1-NN method, but the first three series have almost the same error (between 0.19% and 0.20%). The 1-NN method classifies data with a better error of 0.10%; time required to obtain results with this method of classification is 10894.8 s (about 181.5 min) with the use of the testing computer. The bubble algorithm is faster. During the first measurement series (the testing set contained 49394 elements and the learning set contained 444627 elements), the bubble algorithm obtained the classification result of 0.19%. The average computation time in the learning phase of this test series is 7056.3s (117.6 min) and for the classification phase - 689.1s (11.5 min), i.e. in total over 3149.4s (52.5 min) better than the 1-NN method (this is an acceleration by above 29%).

The second series (the testing set contained 98795 elements and the learning set contained 395226 elements) gave a little worse classification quality, that is to say 0.20%. The time required to obtain this result was the average of 5240.6s (87.3 min) for the learning phase and 1194.2s (19.9 min) for the classification

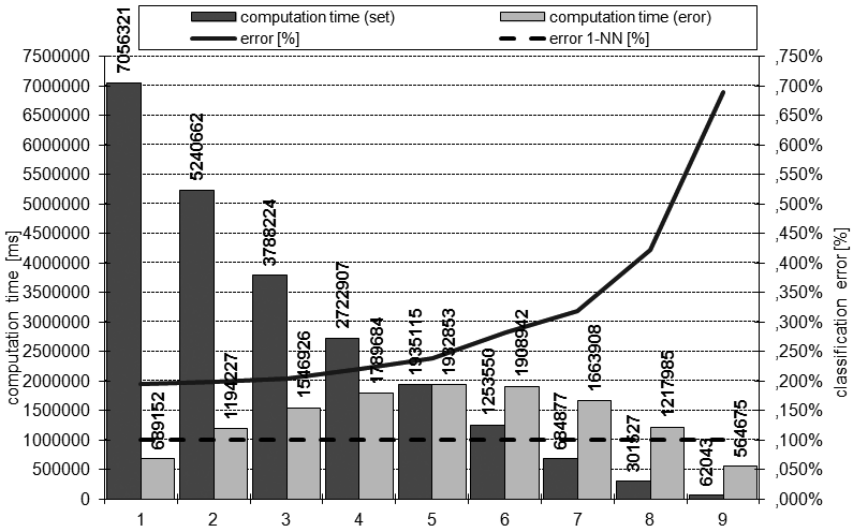


Fig. 4. The presentation of the average values of the algorithms operation time, error calculation time for the Network set

phase. In total, these computations took the author 4459,9s (74,2 min) less compared to the 1-NN rule. This is an acceleration by 40%.

The third series (the testing set contained 148198 elements and the learning set contained 345823 elements) gave comparable classification quality to the former series. This result was 0.20%, too. The learning phase computation time was 3788.2s (63.1 min) and the classification phase computation time was 1546.9s (25,8 min). In total, these computations took the author 5335.1s (88.9 min less than in case of the 1-NN rule). This is an acceleration by 51%.

The next measurement series gave deteriorated classification results. In the practical application this increase in the number of incorrect decisions disqualifies the method; however, a good classification quality and the considerable acceleration of computations in the first few series suggest that maintaining correct proportions between the testing set and the learning set may be beneficial. Tests performed with the use of the remaining sets revealed that, apart from the proportions, the number of objects is also important (in each case the first few series were the best concerning classification quality).

6 Summary

In the article a new method was presented for the condensation of the reference set, i.e. the bubble classification algorithm. Its construction was inspired by a clear lack of algorithms which could manage a large number of computations performed during the processing of a very large set. This algorithm obtained a

considerable acceleration of computations (50%) connected with the classification of the large NETWORK set, simultaneously giving minimally poor classification results. Efforts should be taken to improve the presented solution. Further works are going to be focused on the optimum division into the learning part and the testing part so as to obtain the highest possible acceleration of computations with unchanged or even better classification quality.

References

1. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. John Wiley & Sons, Inc. (2001)
2. Fix, E., Hodges Jr., J.L.: Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties, Project 21-49-004, Report No. 4, US AF School of Aviation Medicine, Randolph Field, Tex., pp. 261–279 (1951)
3. Sánchez, J.S., Pla, F., Ferri, F.J.: On the use of neighbourhood-based non-parametric classifiers. Pattern Recognition Letters 18(11-13), 1179–1186 (1997)
4. Sierra, B., Larrañaga, P., Inza, I.: K-Diplomatic Nearest Neighbour: giving equal chance to all existing classes. Journal of Artificial Intelligence Research (2000)
5. Hort, R.D., Fukunaga, K.: A new nearest neighbour distance measure. In: 5th IEEE International Conference on Pattern Recognition, Miami Beach, Florida, USA, pp. 81–86 (1980)
6. Short, R.D., Fukunaga, K.: The optimal distance measure for nearest neighbour classification. IEEE Transactions on Information Theory IT-27, 622–627 (1981)
7. Kuncheva, L.I.: Reducing the Computational Demand of the Nearest Neighbor Classifier, School of Informatics. In: Symposium on Computing 2001, Aberystwyth, UK, pp. 61–64 (2001)
8. Jóźwik, A., Chmielewski, L., Cudny, W., Skłodowski, M.: A 1-NN preclassifier for fuzzy k-NN rule. In: 13th International Conference on Pattern Recognition, Vienna, Austria, vol. IV(D), pp. 234–238 (1996)
9. Keller, J.M., Gray, M.R., Givens Jr., J.A.: A Fuzzy k-Nearest Neighbor Algorithm. IEEE Transactions on Systems, Man, and Cybernetics SMC-15(4), 580–585 (1985)
10. Jóźwik, A.: A learning scheme for a fuzzy k-NN rule. Pattern Recognition Letters 1(5-6), 287–289 (1983)
11. Lee, W., Stolfo, S.J., Mok, K.W.: Mining in a data-flow environment: Experience in network intrusion detection. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 1999) (August 1999)
12. Stolfo, S.J., Fan, W., Lee, W., Prodromidis, A., Chan, P.K.: Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project. Computer Science Department Columbia University (1999)

Author Index

Andrysiak Tomasz 255

Balachandran Kartheepan 263
Bobulski Janusz 3
Boldak Cezary 187

Cacko Arkadiusz 13
Chabrier Sébastien 95
Choraś Michał 285
Choraś Ryszard S. 21
Cierniak Robert 53
Czapiewski Piotr 29

Dobosz Piotr 59

Fabijańska Anna 45
Forczmański Paweł 29, 37
Frejlichowski Dariusz 29

Gabillon Alban 95
Gocławski Jarosław 45
Gronkowska–Serafin Jolanta 169, 179

Hanczewski Sławomir 269, 277
Hołubowicz Witold 285
Hofman Radosław 29

Iwanowski Marcin 13, 245

Kędzierska Marta 269
Knas Michał 53
Knop Michał 59
Kowal Marek 67
Kozik Rafał 285
Kraft Marek 203
Kretowski Marek 187
Kuczyński Karol 293

Lech Piotr 145, 301
Loesdau Martin 95
Luszczkiewicz-Piatek Maria 75, 85

Machala Bernadetta 103
Malik Krystyna 103
Marchewka Adam 237
Marcininiak Andrzej 67
Mazurek Przemysław 111, 119, 169
Monczak Roman 67

Nacereddine Nafaa 127

Obuchowicz Andrzej 67
Okarma Krzysztof 137, 145
Olszewski Ireneusz 309
Onderka Zdzisław 153

Pérez de Prado Rocío 237
Półchłopek Dariusz 153
Pedersen Jens Myrup 263
Pellerin Michael 293
Piórkowski Adam 169, 179
Piechowiak Maciej 269
Pieniążek Paweł 161

Reska Daniel 187
Reszewicz Szymon 245

Słowik Wojciech 277
Saganowski Łukasz 255
Salski Bartłomiej 245
Sarwas Grzegorz 195
Schmidt Adam 203
Sierszeń Artur 319
Skoneczny Sławomir 195
Smiatacz Maciej 211
Smołka Bogdan 85, 103
Stęgierski Rafał 293
Suszyński Waldemar 293
Szczepański Marek 219
Szwoch Mariusz 227
Szwoch Wioleta 227

Tecław Mateusz 145

Węgliński Tomasz 45

Zdrojewski Jarosław 237
Zieliński Bartłomiej 245
Ziou Djemel 127