ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING 184

Ryszard S. Choraś (Ed.)

# Image Processing and Communications Challenges 4

Springer

# Advances in Intelligent Systems and Computing 184

**Editor-in-Chief**

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

For further volumes:
http://www.springer.com/series/11156

Ryszard S. Choraś (Ed.)

# Image Processing and Communications Challenges 4

Springer

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

ISSN 2194-5357 e-ISSN 2194-5365
ISBN 978-3-642-32383-6 e-ISBN 978-3-642-32384-3
DOI 10.1007/978-3-642-32384-3
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012943845



Printed on acid-free paper

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

# Preface

This textbook collects a series of research papers in the area of Image Processing and Communications which not only introduce a summary of current technology but also give an outlook of potential feature problems in this area.

Image Processing and Communications have undergone an impressive development. Recent evolutions in this area have led to a pervasive spread in many areas of human life and have become such a critical component in contemporary science and technology.

The book is divided into two parts. The first part contains recent research results in image processing, whilst the second part contains recent research results in communications.This textbook collects 36 papers from 9 countries and is the result of the Image Processing and Communications 2012 Conference.

I trust and expect this book to be lasting value to the IT community.

# Contents

# Part I

# Image Processing

# An Approach to Automatic Detection and Extraction of Regions of Interest in Still Images

Dariusz Frejlichowski and Kamil Grzegorzewicz

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

**Summary.** Localisation and extraction of a salient region in an image constitutes an important task in computer vision. It is a very challenging problem due to its ill-posed nature. Furthermore, it is not always evident even for humans. In this paper we present an approach for automatic detection and extraction of salient regions. The main core of the algorithm is based on Ullman and Koch's model of the bottom-up attention, however their approach usually predicts only the place of a salient region. Therefore, in order to improve the achieved results a segmentation method based on k-means has been adopted. In addition, a few heuristic limitations are applied to the previous segmentation method in order to prevent it from creating inadequate segments. We show that our approach can render quite precise results in localisation and extraction tasks, taking into account the human point of view.

## 1 Introduction

One of the most important tasks in digital image analysis is the automatic extraction of object(s) of interest [1], often comprehended as region of interest (ROI). Since more and more data is available in the global network, the importance of this problem is constantly growing. There are several reasons for finding solutions to the problem of extracting objects in an image. The most important one is the application of such methods in the field of content based image retrieval (CBIR), which can lead to the improvement of the results. Another reason is the reduction of high computational complexity in the further processing of an image [2]. For instance, if we focus on a small region of an image, we can improve the expected results as a result of the application of more computationally demanding algorithms. It can be successfully used in various domains such as image compression, digital watermarking, or graphic designing [2, 3, 4].

Obviously, before proceeding with the extraction, the region of interest on an image, often denoted in literature as a salient region, has to be found. Finding this region, which meats particular expectations, is not a trivial task.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 3–10.
springerlink.com 

In many cases it needs to be simplified. Recently, many models attempting to manage this task were brought forward. They can be divided into two main categories based on the level of information assured to images. The first one is the bottom-up attention, which takes into account only low-level features of an image, such as colour, texture and orientation in order to predict the region of interest. There are many examples of approaches belonging to this group [5]. One of the most known approaches is the human attention model devised by Koch and Ullman [6], which is quite plausible in reality. It was applied in works presented in several papers [1, 3, 7, 8], focusing on finding an object of interest in the way that a human would do it. The second category is the top-down attention, which demands additional information given by the human source, often in the form of expert knowledge [9]. One of the examples is the category-object detection [10, 11, 12, 13], currently one of the most dynamically developed domains of computer vision. It is based on the assumption that we can detect only a couple of previously defined object categories. Another example of the top-down model is matting, which is popular in photography. It is a semi-automatic way to detect and extract an appropriate object in an image. It often requires some preparation by a human, but it provides very promising result [14].

In this paper we focus on finding the location and shape of ROI, basing on the model of bottom-up attention by means of the Koch's model. Additionally, a method of image segmentation will be used. The goal is to improve the shape of the extracted region, in order to interpret it in a human way.

## 2 Salient Detection and ROI Extraction

The extraction of appropriate regions of an image can be formulated by means of the binary labelling problem [9]. It means that all of the points belonging to the object(s) of interest are in the same segment and the rest of the image is regarded as the background.

### 2.1 Architecture of the Attention Model

In order to detect primary saliency regions of an image we have used the bottom-up attention model proposed by Koch and Ullman [6, 7], which is based on saliency maps. This model predicates, through special maps, the importance of each image pixel that is salient for humans. Results are usually presented in the form of an intensity map (the saliency map), where the lighter the point on the map, the more important it is.

The processing of an image can be divided into several stages related to the appropriate feature maps (this idea is pictorially presented in Fig. 1). The input image $I$ is down-sampled to half of its previous size. This operation is repeated until the width and height of each output image reaches 1. Down-sampling is performed by means of convolution with a Gaussian filter,

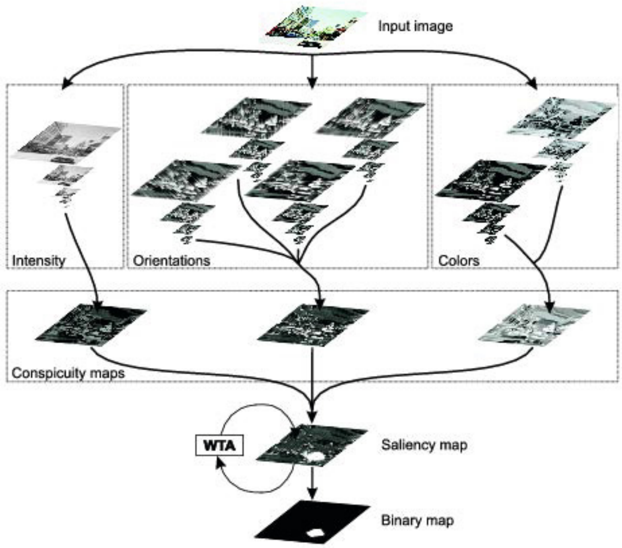



**Fig. 1** An illustration of the visual attention model based on the model proposed by Koch and Ullman ([6]). The description is provided in the main text.

and decimation by a factor of two. The results of this computation are often referred to as the dyadic Gaussian pyramid.

The Gaussian pyramid is subject to several operations. Firstly, each map is brought down to an intensity map by the simple process of averaging out pixel colours, which is defined as ([7]):

$$\mathcal{M}_I = \frac{r+g+b}{3}. \tag{1}$$

The colours of an image are represented by $r$, $g$ and $b$, which stand for red, green and blue channels, respectively. From the intensity maps we formulate additional maps for colours and orientations.

Every level of the intensity maps is decomposed into maps of red-green and blue-yellow, which humans denote as opponent colours that cannot be perceived at the same time. Those maps can be obtained by means of the following formulas ([7]):

$$\mathcal{M}_{RG} = \frac{r-g}{\max(r,g,b)}, \qquad \mathcal{M}_{BY} = \frac{b-\min(r,g)}{\max(r,g,b)}. \tag{2}$$

Another group of maps based on the intensity Gaussian pyramid is created by means of a set of appropriately and variously oriented Gabor filters. Every level of the pyramid is convoluted by Gabor filters ([7]):

$$\mathcal{G}_{\psi}(x,y,\theta) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\delta^2}\right) + \cos\left(2\pi\frac{x'}{\lambda}+\psi\right), \tag{3}$$

where: $\gamma$ is the aspect ratio, $\delta$ denotes the standard deviation, $\lambda$ is a wavelength and $\psi$ is a phase. Coordinates ($x'$,$y'$) are obtained by simple trigonometric transformations with respect to the orientation $\theta$. The convolution with Gabor filters is treated as spatial filtering in primary visual cortex.

Having these three kinds of maps (feature maps) we can begin to compute the human plausible operation. The first one is the center-surround receptive fields, which are simulated by the across-scale subtraction ($\ominus$) between two maps of the pyramid: center ($c$) and surround ($s$) levels ([7]):

$$\mathcal{F}_{l,c,s} = \mathcal{N}(|\mathcal{M}_l(c) \ominus \mathcal{M}_l(s)|) \qquad \forall l \in L = L_I \cup L_C \cup L_O, \tag{4}$$

where: $\mathcal{M}_l$ denotes the type of feature map and $\mathcal{N}$ is an iterative, non-linear normalisation operation caused by different sizes of maps.

The second human plausible operation is the center-surround combination exercising across-scale addition ($\bigoplus$):

$$\bar{\mathcal{F}} = \mathcal{N}\left(\bigoplus_{c=2}^{4}\bigoplus_{s=c+3}^{c+4}\mathcal{F}_{l,c,s}\right) \qquad \forall l \in L. \tag{5}$$

Once again, after the addition, we have to perform the normalisation $\mathcal{N}$.

After the subtraction and addition of feature maps, all of them are once more summed and normalised, firstly to conspicuity maps, which contain general information about the intensity, colours and orientations of the input image. Secondly, they form one saliency map. The location of an attended region in an image is computed on a saliency map by means of the winner-take-all (WTA) neural network, which is described in details in [7]. The region of interest is denoted by the conspicuity map, which emphasizes the activity in the most salient location.

### 2.2 Segmentation

Regardless of ROI's localisation, segmentation is also performed on the input image. For the input images the algorithm operating in the colour domain proposed by Athi Narayanan S. [15] has been applied. The main assumption underlying this method is the similarity of neighbouring pixels — a new value for every point of an image is computed in the window surrounding that pixel ([15]):

$$x_{new} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=0}^{2}\frac{(x_{ij}/div)k}{bins}, \tag{6}$$

where $x_{ij}$ represents a pixel in the window, $N$ stands for its size, $div$ and $bins$ are experimentally determined constants. This operation is called 'histogram based windowing process' and provides us with a coarse image representation. The new image values stand for the arguments for $k$-means clustering, based on the Euclidean metric, with a pre-defined number of clusters. Later, the process of clusterization is conducted in order to obtain the segmented image.

### 2.3 Appropriate ROI Estimation

In order to determine the ROI in a human way, knowledge of past experience is usually required. However, because we are only using the low-level features in an image we have to estimate the shape and predict the location of objects.

In this paper the prediction of a salient region is based on the attention model, in which the segmentation algorithm brings the estimation of a more appropriate ROI shape. It is important to perform the extension of a shape to object(s) of interest, because we often gain a so-called proto-object ([7]): "which is described as volatile units of visual information that can be bound into a coherent and stable object when accessed by focused attention". It does not focus on the precise boundaries of an object.

Basing on the previous selection of pixels, which belong to the ROI, we took all segments of an image overlapping it in the spatial domain. This process is based on the assumption that neighbouring pixels are relatively similar to each other and they belong to the same object.

There are some limitations in the process of selecting segments. The first one refers to the spatial size of extracted regions of an image — it can not be larger than half of the image's size. If this condition is fulfilled, the morphologic operation of closing can be used in order to join pixels with the same attributes for better segment integration. Another limitation concerns the spatial distribution of segments which aims to eliminate regions characterised by high spatial variation. The side effect of this operation allows to create shapes larger than half of the size of an image, since it is based on new, more homogeneous segments.

## 3 Experimental Evaluation of the Approach

The evaluation of the proposed approach was conducted on the image database introduced in [16]. The authors categorised outdoor images into eight groups: city, coast, country, forest, highway, mountain, street and tall building. The images have the same size — $256 \times 256$ pixels. They were labeled and the average number of objects on an image was equal to 11. Some examples are provided in Fig. 2.

The efficiency of the presented method could only be evaluated by humans. Therefore, the results of the approach were carefully examined and judged by four persons. Each of them was asked to evaluate 250 processed images. An image was examined by more than one person, due to the subjectivity of this

**Fig. 2** Exemplary images used in the experiments divided into eight groups

**Fig. 3** Examples of tested images (on the left — an input image, on the right — the result; black regions denote irrelevant pixels). The first two rows contain images with respondent positive opinion, last row with negative.

process. It is obvious that such factors as education, age, interests, hobbies, etc. can significantly influence the results provided by a particular person.

The experimental results are presented in Table 1, separately for each class. Every respondent was asked to make a binary decision — whether in his/her opinion a processed image contains salient regions or not. Afterwards, all the answers were summed up and divided by the number of test images.

**Table 1** Results of the proposed approach, based on evaluation performed by humans

| Category | ROI accordance with human judgements |
|---|---|
| City | 96% |
| Coast | 95% |
| Country | 86% |
| Forest | 89% |
| Highway | 97% |
| Mountain | 90% |
| Street | 91% |
| Tall building | 77% |
| Overall | 90% |

The results presented in Table 1 vary according to particular categories. The worst results were achieved for tall buildings. This reveals one of the most important limitations of the proposed approach, namely that the method generates false results in the case of images containing big spatial objects with very low pixel similarity. An example of such image is presented in Fig. 3 (located on the right, in the bottom row). For other categories the results were considerably better.

## 4 Conclusions

In the paper we have presented a method for automatic detection and extraction of salient regions, performed similarly to the process of human understanding. It is very important to have information about ROI on images, as it may significantly reduce unnecessary computational operations in further processing. In many cases indicating only the location containing salient objects is insufficient. It has to be enhanced. In order to obtain better results we have introduced a segmentation method in order to improve a region's shape. Furthermore, we have added a simple heuristic in order to rule out achieving undesirable regions according to the human point of view.

The experiments with the proposed approach gave very promising results. Additionally, in many cases the achieved image contained more than one object of potential interest. However, in future works, a more objective evaluation of the proposed method should be performed, e.g. using some 'ground-truth data'.

## References


1. Anouncia, S.M., Joseph, J.G.: An Approach for Automated Object Recognition and Extraction from Images — a Study. Journal of Computing and Information Technology 17(4), 359–370 (2009)
2. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to Predcit Where Humans Look. In: IEEE Int. Conf. on Computer Vission, ICCV 2009 (2009)
3. Ko, B.C., Nam, J.Y.: Object-of-Interest Image Segmentation Based on Human Attention and Semantic Region Clustering. Optical Society of America 23(10), 2462–2470 (2006)
4. DeCarlo, D., Santella, A.: Stylization and Abstraction of Photographs. ACM Trans. Graph. 21(3), 769–776 (2002)
5. Jain, A.K., Ratha, N.K., Lakshmanan, S.: Object Detection Using Gabor Filters. Pattern Recognition 30(2), 295–309 (1997)
6. Koch, C., Ullman, S.: Shifts in Selective Visual-Attention — Towards the Underlying Neural Circuitry. Human Neurobiology 4, 219–277 (1985)
7. Walther, D., Koch, C.: Modeling Attention to Salient Proto-objects. Neural Networks 19(9), 1395–1407 (2006)
8. Zhang, J., Zhuo, L., Shen, L.: Region of Interest Extraction Based on Visual Attention Model and Watershed Segmentation. In: IEEE International Conference on Neural Networks & Signal Processing, pp. 375–378 (2008)
9. Liu, T., Sun, J., Zheng, N.N., Tang, X., Shum, H.Y.: Learning to Detect A Salient Object. In: IEEE Conf. on Computer and Vision Pattern Recognition, pp. 1–8 (2007)
10. Larlus, D., Jurie, F.: Combining Appearance Models and Markov Random Fields for Category Level Object Segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1–7 (2008)
11. Aldavaret, D., Ramisa, A., Toledo, R., López de Mántaras, R.: Fast and Robust Object Segmentation with the Integral Linear Classifier. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1046–1053 (2010)
12. Murphy, K., Torralba, A., Eaton, D., Freeman, W.: Object Detection and Localization Using Local and Global Features. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds.) Toward Category-Level Object Recognition. LNCS, vol. 4170, pp. 382–400. Springer, Heidelberg (2006)
13. Marszalek, M., Schimd, C.: Semantic Hierarchies for Visual Object Recognition. In: IEEE Conf. on Computer Vision and Pattern Recogniton, pp. 1–7 (2007)
14. Wang, J., Cohen, M.F.: Image and Video Matting: a Survey. Foundations and Trends in Computer Graphics and Vision 3(2), 97–175 (2007)
15. Athi Narayanan, S.: Homepage, http://sites.google.com/site/athisnarayanan/
16. Oliva, A., Torralba, A.: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. Int. J. Comput. Vision 42(3), 145–175 (2001)

# Decision Tree Based Approach to Craquelure Identification in Old Paintings

Joanna Gancarczyk

University of Bielsko-Biala, Willowa 2, 43-309, Bielsko-Biala, Poland
jgan@ath.bielsko.pl

**Summary.** In this paper an attempt has been made to develop a decision tree classification based algorithm for craquelure identification in old paintings. Craquelure can be an important element in judging authenticity, artist's workshop as well as for monitoring the environmental influence on the condition of the painting. Systematic observation of craquelure will help to build a better platform for conservators to identify cause of damage, thus a proper tool for precise detection of the pattern is needed. However, the complex nature of the craquelure is a reason why an automatic detection algorithm is not always possible to implement. The result presented in this work is an extension of known semi-automatic technique based on a region growing algorithm. The novel approach is to apply a decision tree based pixel segmentation method to indicate the start points of craquelure pattern. This, in particular applications may improve significantly the overall effectiveness of the algorithm.[1]

## 1 Introduction

Craquelure is a pattern of cracks that appears in a painting during the process of aging. Cracks in the paint layer grow as the canvas or wood support of the painting moves in response to changes of humidity and temperature. Every layer of a painting has it's own distinctive mechanical behaviour, and therefore, every layer contributes in it's own way to the formation of craquelure [6], [5]. A classification of features for cracks on paintings was made by Stout [6]. Apart from the local features of individual cracks, like the depth or the smoothness of its edges, research was also performed towards defining characteristics of craquelure patterns. In [4] Bucklow proposed a classification of crack patterns of paintings into four different categories, representing four paint traditions: Italian fourteenth/fifteenth-century paintings on panel, Flemish fifteenth/sixteenth-century paintings on panel, Dutch seventeenth-century paintings on canvas and French eighteenth-century paintings on canvas. De Willingen in [5] reported a detailed study of molecular and mechanical

[1] This work was supported by NCN (National Science Centre) under grant no. 6593/B/T02/2011/40.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 11–17.
springerlink.com 

issues concerning generation of stress between paint layers and their influence on formation of crack pattern. Figure 1 is an example of a crack pattern on an oil on canvas 19th century painting from the collection of the National Museum in Krakow.

**Fig. 1** Crack pattern on the detail of 19th century oil on canvas painting (Rafał Hadziewicz, "Portrait of Wentzl")

Ridge and valley structure detection, of which craquelure pattern identification is a good example has been a matter of high interest among researchers, mostly for its potentially useful contribution to a variety of applications. The results presented in this work have a much wider set of applications than just the analysis of paintings. Many images contain similar patterns, such as biological images of veins and tissues, images of fingerprints and multi-spectral satellite photos of rivers or roads. All these examples are current areas of research in the modelling and classification fields.

## 2 Virtual Restoration

Because of the increasing availability and effectiveness of digital imaging tools a significant step has been observed towards popularization computer vision technique in the art analysis ground. In [9] Stork indicates brush stroke and craquelure identification, dewarping, perspective and lighting analysis as most exploited and promising up to date research areas. Other authors mention also recomposition of fragments, virtual restoration and lacuna filling (see [10], [12], [13] and [14] for summary analysis).

With image processing tools it is possible to entirely remove cracks by means of interpolation techniques ([7], [8]). An algorithm for crack removal is usually a two-step procedure: first, the cracks have to be detected; next, the selected cracks are filled in. The crack selection step can be semi-automatic, or automatic. In the first case, the aid of the user is requested to let him choose a starting point of the crack, from which it is possible to start an iterative process able to identify the whole crack by assuming the hypothesis that a crack is darker than the background and it is characterised by a rather elongated structure ([11]). In automatic crack selection, cracks are identified by means of a proper filter, like Gabor filters, or a morphological filter called top-hat transform ([1], [2], [3]). However, with this approach not only cracks, but also brush strokes and other texture characteristic could be detected. This problem can be solved by discriminating brush strokes and cracks on the basis of shape, hue or saturation values or thickness and favoured orientation [13].

## 3 Decision Tree Representation

A decision tree is built from a training set, which consists of feature vectors, each of which is constructed by a set of attributes and a class label. Attributes are descriptors containing information about the object. Each attribute may have either ordered or unordered values, of discrete or continuous type. Several methods have been proposed to construct decision trees, the initial being based on CART and C4.5 algorithms. These algorithms generally use the recursive-partitioning algorithm, and its input requires a set of training examples, a splitting rule, and a stopping rule. Partitioning of the tree is determined by the optimization of a given parameter, like Gini index or entropy. The splitting procedure is recursively repeated for each subset until no more partioning is possible. Stopping rules for the algorithm are applied depending on the application, but there are some general solutions defined. One stopping rule is to test for the purity of a node. Another stopping rule is by looking at the depth of the node, defined by the length of the path from the root to that node. Third method is the example size. If the number of examples at a node is below a certain threshold, then splitting is not allowed.

The Microsoft Decision Trees algorithm uses different methods to compute the best tree. The method used depends on the task, which can be linear regression, classification, or association analysis. A single model can contain multiple trees for different predictable attributes. Moreover, each tree can contain multiple branches, depending on how many attributes and values there are in the data. The shape and depth of the tree built in a particular model depends on the scoring method and other parameters that were used. Changes in the parameters can also affect where the nodes split. A tree is built by determining the correlations between an input and the targeted outcome. After all the attributes have been correlated, the algorithm identifies the single attribute that most cleanly separates the outcomes. This point of the best separation is measured by using an equation that calculates information gain. The attribute that has the best score for information gain is used to divide the cases into subsets, which are then recursively analysed by the same process, until the tree cannot be split any more [15].

# 4 Decision Tree Based Approach to Craquelure Separation

## 4.1 Training Data

Definition of a training set for further construction of a decision tree is based on a manual selection of craquelure pixels in a small area selected as a representative region of the whole image. This region should be chosen according to the colour, brightness and texture, since the best results will be obtained in the final step of the algorithm for regions similar to the training set. In case the painting is not consistent, that means the parts differ from one another significantly a few training areas may be chosen concurrently. See Figure 2 to observe the training points selection for the analysed fragment of a painting.

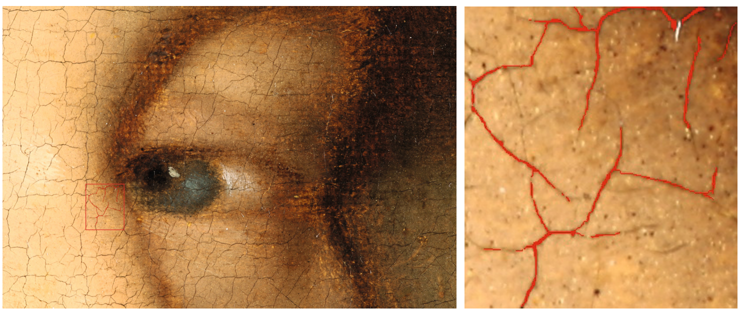


**Fig. 2** Training set definition for a decision tree

Once the training set is defined and craquelure pixels selected a binary mask is created for the training area. See Figure 3(left) to compare.

## 4.2 Feature Vector Generation

For the definition of feature vectors upon pixels of the image several values are computed. The basic set of parameters consists of the R, G and B channel values and a grayscale value for each pixel. Alternatively, the L*a*b space values may be chosen. It's verified, that morphological operations and spatial filtering give relatively good results for a detection of ridge-and-valley structures, like craquelure pattern analysed in this work. Therefore another set of values was computed on basis of the grayscacale image, that is image opening, closing, median filtering and the difference between median filtered image and the original. The operations were done by MATLAB functions `imopen`, `imclose` and `medfilt2` respectively with a square structuring element of size 15.

### 4.3 Setting Up a Decision Tree

The set of features obtained for each analyzed pixel was exported to relational database management system for further processing. For this stage MS SQL Server 2008R2 was used with the Business Intelligence Development Studio features. A mining structure was defined with application of the Microsoft Decision Trees model. A source table consisted of feature vectors for each pixel in the training set area and class labels. There are two class labels 1 and 0 for craquelure and non-craquelure pixels respectively.

## 5 Results

The mining model applied in the crack selection task distinguished three significant classes of pixels that may be assumed to be craquelure pixels. The classes are as follows: (1) - difference between gayscale value and median filtered image value ranging from 26 to 39 and red value less than 100; (2) - difference ranging from 39 to 52, median filtering less than 162 and red value less than 100; (3) - difference ranging from 52 to 104 and grayscale value less than 103. That means only four parameters appeared to me meaningful while performing the computations. Figure 3 presents the mask generated in the training data preparation step compared with a new mask based upon the rules obtained by building the decision tree. The latter is intended to be an initial mask for craquelure pattern detection by means of e region growing algorithm. Several missclassification may be observed in the right side picture. This is partially due to the character of the image - no smoothing nor noise reduction process was applied before the actual computation. Also the set of parameters of vector features might be improved. This would be a subject of further research.

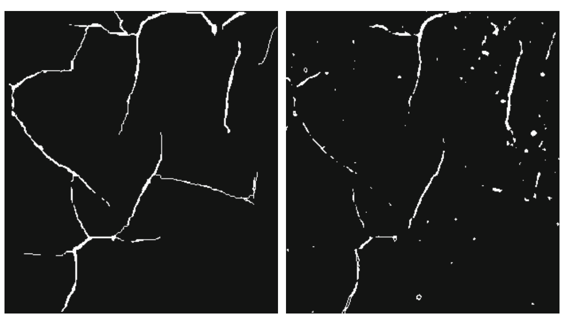


**Fig. 3** Mask and the computed initial set of craquelure pixels for the same area

**Fig. 4** Initial set of craquelure pixels for the whole image (negative)

## 6 Conclusions and Further Work

A new technique of craqelure segmentation from old paintings was presented and some possible application areas proposed. The novel approach is based on application of a decision tree segmentation method to distinguish an initial set of craquelure pixels instead of pointing them manually. The effectiveness of this method depends of the ratio of work put into selecting crack pixels for the training set of a decision tree to the work, that would by performed in the manual method. That means, the best results, comparing to the manual method, are achieved when the crack pattern is inconsistent (non solid) and the background relatively homogeneous. The attributes of a decision tree are based on the colour of particular pixels, grayscale value and grayscale value after median filtering. Further work will concern better adjustment of the feature vector to obtain more suitable classification results.

**Acknowledgement.** Author would like to thank Mrs. Joanna Sobczyk and the Laboratory of Analysis and Nondestructive Investigation of Heritage Objects of the National Museum in Krakow for substantial support and making accessible high resolution images of paintings from the collection of Rafał Hadziewicz works.

## References

1. Abas, F.S.: Analysis of Craquelure Patterns for Content-Based Retrieval. PhD Thesis, University of Southampton, Southampton (2004)
2. Abas, F.S., Martinez, K.: Classification of painting cracks for content-based analysis. In: IST/SPIE's 15th Annual Symp. Electronic Imaging, Santa Clara, California, USA (2003)

3. Abas, F.S., Martinez, K.: Craquelure analysis for content-based retrieval. In: Proc. of 14th Int. Conf. on Dig. Sig. Proc., Santorini, Greece, pp. 111–114 (2002)
4. Bucklow, S.L.: A sylometric analysis of Craquelure. Computers and the Humanities 31, 503–521 (1998)
5. De Willigen, P.: A Mathematical Study on Craquelure and other Mechanical Damage in Paintings. Delft University Press, Delft (1999)
6. Stout, G.L.: A trial index of laminal disruption. JAIC 17(1), Article 3, 17–26 (1977)
7. Hanbury, A., Kammerer, P., Zolda, E.: Painting crack elimination using viscous morphological reconstruction. In: Proc. ICIAP 2003, Mantova, Italy (2003)
8. Gupta, A., Khandelwal, V., Gupta, A., Srivastava, M.C.: Image processing methods for the restoration of digitized paintings. Thammasat Int. J. Sc. Tech. 13(3), 66–72 (2008)
9. Stork, D.G.: Computer Vision and Computer Graphics Analysis of Paintings and Drawings: An Introduction to the Literature. In: Jiang, X., Petkov, N. (eds.) CAIP 2009. LNCS, vol. 5702, pp. 9–24. Springer, Heidelberg (2009)
10. Cappelllini, V., Barni, M., Corsini, M., de Rosa, A., Piva, A.: ArtShop: an art-oriented image-processing tool for cultural heritage applications. J. Visual Comput. Animat. 14, 149–158 (2003)
11. Barni, M., Bartolini, F., Cappellini, V.: Image processing for virtual restoration of artworks. IEEE Multimedia 7(2), 34–37 (2000)
12. Barni, M., Pelagotti, A., Piva, A.: Image processing for the analysis and conservation of paintings: opportunities and challenges. IEEE Sig. Proc. Mag. 141 (2005)
13. Cappellini, V., Piva, A.: Opportunities and Issues of image processing for cultural heritage applications. In: Proc. EUSIPCO 2006, Florence, Italy (2006)
14. Sobczyk, J., Obara, B., Frączek, P., Sobczyk, J.: Zastosowania analizy obrazu w nieniszczących badaniach obiektów zabytkowych. Wybrane Przykłady, Ochrona Zabytków 2, 69–78 (2006)
15. Microsoft Decision Trees Algorithm Technical Reference, http://msdn.microsoft.com/en-us/library/cc645868

# Computer Vision Method for Detecting Adult-Oriented Content in Images

Adam Marchewka and Rafał Kozik

University of Technology and Life Sciences, Bydgoszcz, Poland
adam.marchewka@utp.edu.pl

**Summary.** As people started using the Internet a huge volume of data has become available on-line. Now it is impossible to efficiently control its content that may be unwanted or inappropriate for children. This topic is even more challenging when it comes to images and video sequences. Therefore, the problem of detecting an adult-oriented content in images still remains unsolved and is widely addressed by many researchers. In this paper we propose a low-level method for filtering such unwanted elements. It is based on a novel approach to skin segmentation that engages Gaussian distributions. The algorithm is self-calibrating and does not require any pre-defined thresholds and parameters, which is a certain disadvantage of many skin segmentation algorithms.

## 1 Introduction

The problem of detecting harmful content in Internet such as pornography is currently one of the most important research topics. It becomes even more challenging when it comes to image and video data. In this paper we propose a low-level method for filtering adult-oriented elements, that is based on skin detector. The proposed algorithm for skin segmentation engages a mixture of Gaussian distributions. The algorithm is self-calibrating and does not require any pre-defined thresholds and parameters, which is a certain disadvantage of many skin segmentation algorithms.

The paper is organized as follows. First, in section 2 some of existing methods for detecting the adult-oriented content are described and discussed. In section 3 the algorithm architecture is shown and explained. The results are presented in section 4. Afterwards final conclusions and plans for future work are presented.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 19–24.
springerlink.com 

## 2 Methods for Detecting Adult-Oriented Content

The methods for adult-oriented image content detection can be classified into three categories:

1. based on the skin color segmentation,
2. based on the texture descriptors,
3. hybrid methods (that combine group 1 and 2).

One of the earliest works on skin color segmentation was presented by Hunke in [1]. The author made assumption that when brightness is normalized, the skin colors distribution for different human races can be efficiently approximated with single Gaussian distribution.

Fleck et al. algorithm described in [11] focuses on detecting naked people with an algorithm involving a skin and geometric properties of skin regions.

The system presented in [12] uses Daubechies wavelets, moment analysis, and histogram indexing in order to provide feature vectors that are used for understanding the content of the image.

More recent approaches such as this one presented by Tang *et al.* in [2] uses a multiple Gaussian distributions with normalized color spaces, namely: $r - g$ and $NC_b - NC_r$. The images are first transformed into the two normalized color spaces. Then the skin color similarities of each image (that has been normalized) pixel are measured in the $r - g$ space and in the $NCb - NCr$ space, respectively. The image with skin region is obtained by combining two similarities.

## 3 Proposed Method

In this section the algorithm for skin segmentation is explained and discussed. The block diagram of the proposed algorithm is presented in Fig.2. It consists of following steps:

1. The large images are downscaled to 128×128 resolution in order to produce smoothed regions and speed-up computation process.
2. The number of color channels for each image is reduced from 3 ($RGB$) to 2 ($Y_r$ and $Y_b$) using formula (1).

$$\begin{cases} I = & 0.299R + 0.587G + 0.114B \\ Y_r = R - I \\ Y_b = B - I \end{cases} \tag{1}$$

3. The 2D histogram of colors in $Y_rY_b$ space is created (Fig. 1a). The samples representing the skin were randomly selected from sample images.
4. The first 5 maximum values in 2D histogram are selected.

5. For each maximum the skin color is approximated by a Gaussian function represented by equation (2).

$$G(Y_r, Y_b) = \frac{1}{2\pi\alpha^2} = e^{-\frac{Y_r^2+Y_b^2}{2\alpha^2}} \tag{2}$$

The $\alpha$ is introduced in order to smooth the skin color region (Fig. 1b).

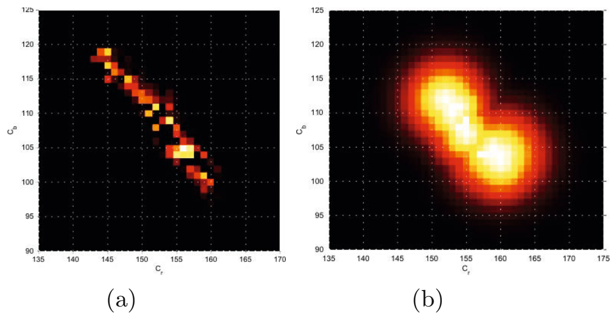



**Fig. 1** Distribution of skin-color pixels in YCbCr space (a) before (b) after multiple Gaussian approximation

6. Using the proposed model for each pixel of the original image $I[x, y]$ a probability $I_p[x, y] = G(I[x, y])$ of skin is assigned.
7. The $I_p[x, y]$ is additionally blurred with median filter in order to produce the $I_f[x, y]$ image.
8. The final probability image, defined as $I'_p[x, y] = I_f[x, y] + I_p[x, y]$ is normalized in order to achieve probability values in $< 0; 1 >$ range.
9. Multiplying the probability image $I_p[x, y]$ with original image $I[x, y]$ a skin region is extracted. Each pixel value in original image is encoded using one byte. Therefore multiplying it with particularly low probability produces 0 value ($I_{mask}[x, y]$).

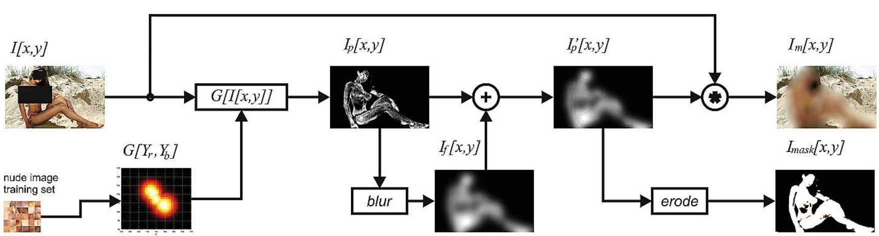



**Fig. 2** Structure of the algorithm

## 4 Results and Future Work

The proposed algorithm was evaluated for database of images containing adult-oriented content. The database was manually created from images available on-line. As it is shown in Fig.3 the method applies strong blur filter in region where skin is detected. It may be notice that non-skin regions remain untouched. Therefore, it easier for user to justify (on his/her own) if the image truly contains inappropriate content.

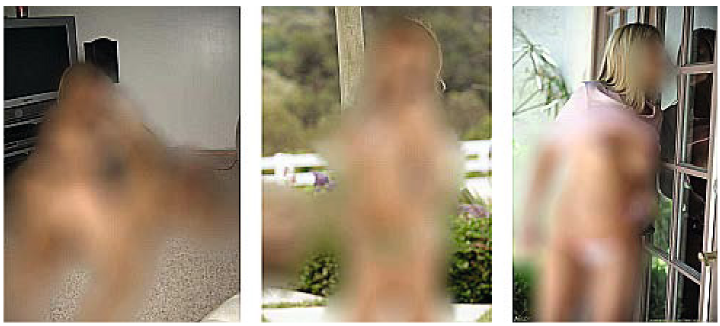


**Fig. 3** Examples of filtered-out adult-oriented content of the image. Strong blur filter is applied in regions identified as skin (used $I_f[x,y]$).

## 5 Conclusions and Future Work

In this paper a low-level detector for adult-oriented image contend was presented. The described algorithm is self-calibrating and does not require any pre-defined threshold and parameters, which is a certain disadvantage of many skin segmentation algorithms. Presented results show that the proposed probabilistic model allows for accurate skin detection and extraction. The proposed method does not analyse the context of image. Therefore, the future work will be focused on providing high-level knowledge that will allow the algorithm to combine the information about skin regions with low-level features descriptors for detecting such object like face or human silhouette. Other possible application areas of described algorithm, that are planed to be investigated, include:

- blurring people faces in video sequences in situation when their anonymity needs to be preserved,

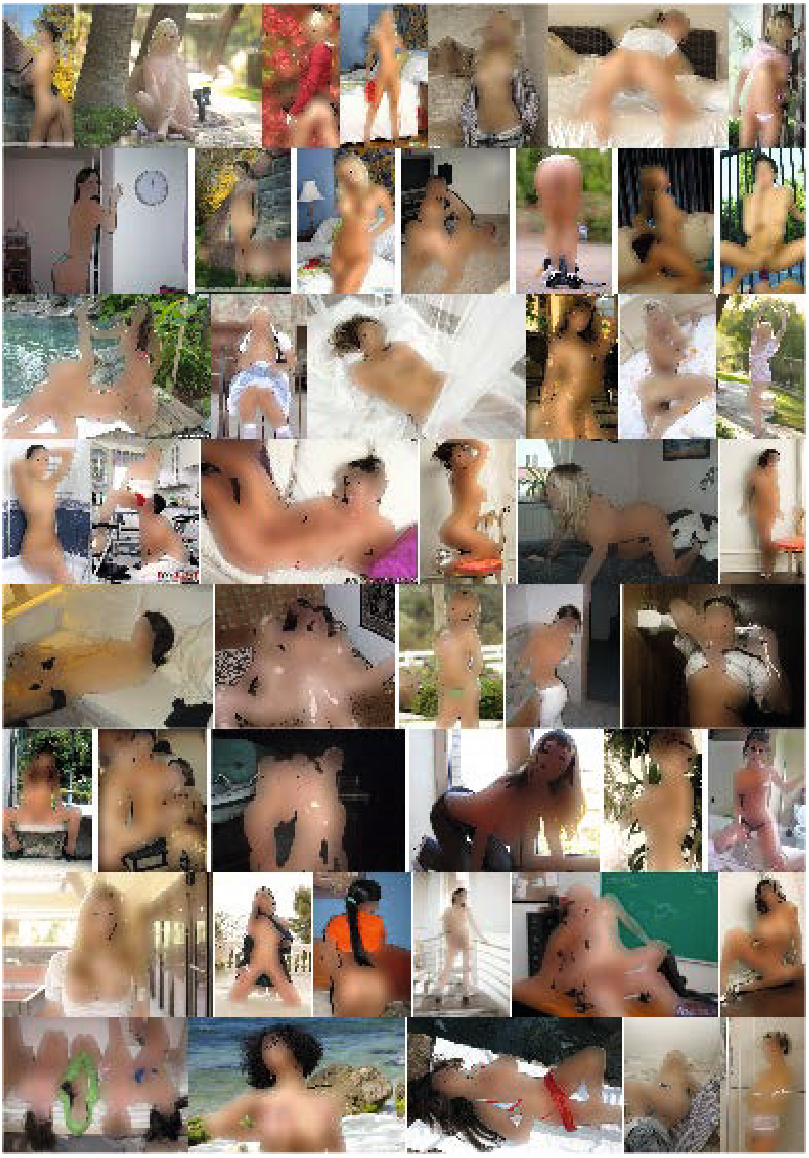


**Fig. 4** Results obtained with proposed algorithm (used $I_{mask}[x, y]$)

- violence detector based on blood color detector with proposed algorithm,
- adult oriented image content detectors deployed on web-browsers or proxy servers.

## References


1. Hunke, H.M.: Locating and tracking of human faces with neural networks. Master's thesis, University of Karlsruhe. Ioffe, S. and Fo (1994)
2. Tang, J.S., Kawato, S., Ohya, J.: Face detection from acomplex background. Research Report, ATR MediaIntegration and Communications Research Laboratories, Kyoto, Japan (2000)
3. Liu, L., Sang, N., Yang, S., Huang, R.: Real-time skin color detection under rapidly changing illumination conditions. IEEE Transactions on Consumer Electronics 57(3), 1295–1302 (2011)
4. Mofaddel, M.A., Sadek, S.: Adult image content filtering: A statistical method based on Multi-Color Skin Modeling. In: 2010 2nd International Conference on Computer Technology and Development, ICCTD, November 2-4, pp. 744–748 (2010)
5. Yin, H., Xu, X., Ye, L.: Big Skin Regions Detection for Adult Image Identification. In: 2011 Workshop on Digital Media and Digital Content Management, DMDCM, May 15-16, pp. 242–247 (2011)
6. Bouirouga, H., Elfkihi Lrit, S., Jilbab, A., Aboutajdine, D.: Recognition of adult video by combining skin detection features with motion information. In: 2011 International Conference on Multimedia Computing and Systems, ICMCS, April 7-9, pp. 1–5 (2011)
7. Wang, X., Zhang, X., Yao, J.: Skin color detection under complex background. In: 2011 International Conference on Mechatronic Science, Electric Engineering and Computer, MEC, August 19-22, pp. 1985–1988 (2011)
8. Huang, L., Xia, T., Zhang, Y., Lin, S.: Human skin detection in images by MSER analysis. In: 2011 18th IEEE International Conference on Image Processing, ICIP, September 11-14, pp. 1257–1260 (2011)
9. Bouirouga, H., El Fkihi, S., Jilbab, A., Bakrim, M.: A comparison of skin detection techniques for objectionable videos. In: 2010 5th International Symposium on I/V Communications and Mobile Network, ISVC, September 30-October 2, pp. 1–4 (2010)
10. Zheng, H., Liu, H., Daoudi, M.: Blocking objectionable images: adult images and harmful symbols. In: 2004 IEEE International Conference on Multimedia and Expo., ICME 2004, June 30-30, vol. 2, pp. 1223–1226 (2004)
11. Fleck, M.M., Forsyth, D.A., Bregler, C.: Finding Naked People. In: Buxton, B.F., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1065, pp. 593–602. Springer, Heidelberg (1996)
12. Wang, J.Z., et al.: System for Screening Objectionable Images. Computer Communications 21(15), 1355–1360 (1998)

# Texture Based Firearm Striations Analysis for Forensics Image Retrieval

Ryszard S. Choraś

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

**Summary.** In this article, we propose a Content-Based Image Retrieval method for some identification in forensic fields. An expert simply places the image as a query on a system, which retrieves the most similar images in the database. Then feature vectors are extracted to compare the query image to another image in the database.

## 1 Introduction

This article highlights the challenges in applying image-recognition technology to forensics applications. The processes of forensics computing can be divided into three main areas:

- Image Capture. The imaging process is fundamental to any computer investigation and in reconstructing a crime scene or a scene of an accident,
- Image Processing. The processing software to extract features of the image,
- Investigation.

Forensics science and biometrics both apply various identification sciences (Figure 1). Biometrics applies to a pre-event situation and chooses which mode of identification will be used. Forensics science, however, applies to post-event situations, reconstructs past criminal events and which mode of identification will be used is not known [1].

Various forensics digital multimedia collections are becoming more common and very extensive. The large number of various forensics images in collections, means that manual processing has become impractical. Retrieving relevant images from these collections will, in many cases, involve visual similarity. Forensics images are stored in large databases and many features have been used to annotate and retrieve images. Typically low-level features such as color, texture and shapes of objects are used as features for image retrieval.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 25–31.
springerlink.com 

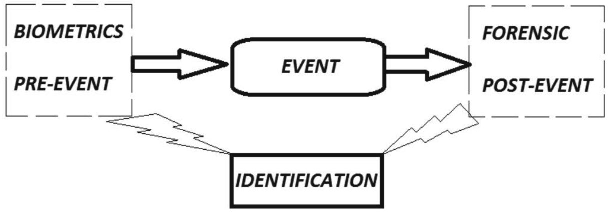



**Fig. 1** Biometrics and forensics

Content-based forensics image retrieval (*CBFIR*), which is based on automatically extracted primitive features such as color, shape, texture, and even the spatial relationships among objects, has been employed and gives practical satisfy results.

It can be described as follows [3]:

1. Index images by their content while image processing operations, Several attributes such as color, texture, shape can be associated with image content.
2. Apply some operators to extract "features" that measure statistical properties of the content attribute. These features can be used to represent the images,
3. Retrieve in a query-by-example fashion.

All firearms, have marks that are unique to a particular firearms manufacturer. Additionally, when a bullet travels through a gun barrel, the bullet's metal gets worn in a unique pattern by the harder metal of the barrel. The land and groove impressions found on the surface of fired bullets (Figure 3a) are used to match bullets with a suspect's weapons [1],[2]. Impressions can be found on cartridge cases, and these can be used to match a cartridge case to a specific firearm. Often the cartridge case is the most important forensics specimen in the identification of weapons, as bullets are commonly deformed by the impact. Firing pin impressions (Figure 3b) , extractor marks (Figure 3c), ejector marks(Figure 3d), chamber marks made by the firearms are also important features used in retrieval firearm. Similarly to bullets, cartridge cases pick individual striation patterns up (Figure 3e). Computer vision techniques have been automated to detect the cartridge/bullets surface roughness [7]. An automated firearms retrieval system ranks the stored images according to certain similarity metrics with a subject's cartridge/bullet.

Marks that are made in a surface as the result of the motion of one surface across the other; can produce striations on one or both surfaces. Striations are most often associated with firearm evidence such as the markings on bullets and cartridge casings.

Retrieval process will attempt to find this unique pattern by following the procedures outlined below.

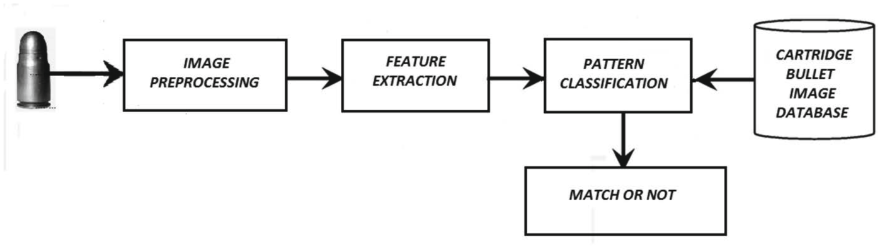
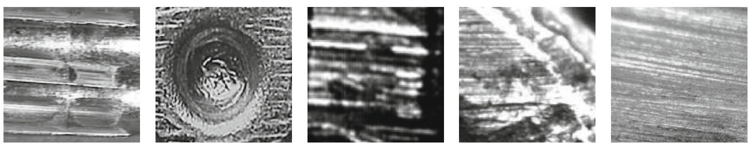



**Fig. 2** CBFIR system



(a) (b) (c) (d) (e)

**Fig. 3** Examples of bullet striation mark

## 2 Feature Extraction

The pictorial information is represented as a function of two variables $(i, j)$. The image in its digital form is usually stored as an two-dimensional array. If $D = N \times N$ is the set of resolutions cells and the digital image $I$ is a function which assigns gray value to each and every resolution cells, i.e. $I : N \times N \longrightarrow G$.

The method of feature extraction is described in this section. Feature extracting plays an important role in retrieval/recognition system. In *CBFIR* systems, some image processing algorithms are used to extract feature vectors that represent striations image properties. To represent striations properties we used the methods known with texture analysis [8].

It provides information in the spatial arrangement of colours or intensities in an image, characterized by the spatial distribution of intensity levels in a neighbourhood and cannot be defined for a point. The co-occurrence matrix was used in extracting statistical features in computing the bullet striation mark. Co-occurrence matrices are second order statistics which describe 2D relations for pairs of grey levels of image pixels.

A co-occurrence matrix represents the distance and angular spatial relationship over an image sub-region of specific size. For a gray-scale image the gray level co-occurrence matrix $(GLCM)$ is calculates how often a pixel with gray-level value $k$ occurs either horizontally, vertically, or diagonally to adjacent pixels with the value $l$ .

Co-occurrence matrix $C_d$ is specified with a displacement vector $d = \{(row, column)\}$. Entry $C_d(k,l)$ indicates how many times a pixel with gray level $k$ is separated from a pixel of gray level $l$ by the displacement vector $d$ (Figure 4).

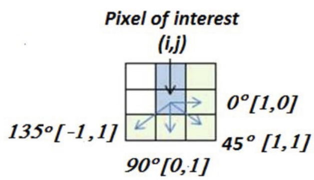



**Fig. 4** Displacement vector

The gray level co-occurrence matrix $C_d(k,l)$ , for distance $d$ can be defined as

$$C_d(k,l) = \sum_{k=0}^{G-1}\sum_{l=0}^{G-1} \begin{cases} 1 \ if \ I(i,j) = k \ \text{ and } \ I(i \pm d, j \pm d) = l \\ 0 \qquad\qquad otherwise \end{cases} \tag{1}$$

Gray level co occurrence matrix $(GLCM)$ is the basis for the Haralick texture features. This matrix is square with dimension $G$, where $G$ is the number of gray levels in the image.

Gray level co-occurrence matrices capture properties of a texture but they are not directly useful for further analysis, such as the comparison of two textures. Numeric features are computed from the co-occurrence matrix that can be used to represent the texture more compactly. One problem with deriving texture measures from co-occurrence matrices is how to choose the displacement vector $d$.

Algorithms for texture analysis are applied to an image in a series of windows of size $w$, each centered on a pixel $(i,j)$ (Figure 5). The value of the resulting statistical measure are assigned to the position $(i,j)$. The feature descriptors computed from co-occurrence matrices are used in texture analysis.

The features of grey level co-occurrence matrices are computed by

- Contrast.

$$Con = \sum_{k=0}^{G}\sum_{l=0}^{G} C_d(k,l)[(k-l)]^2 \tag{2}$$

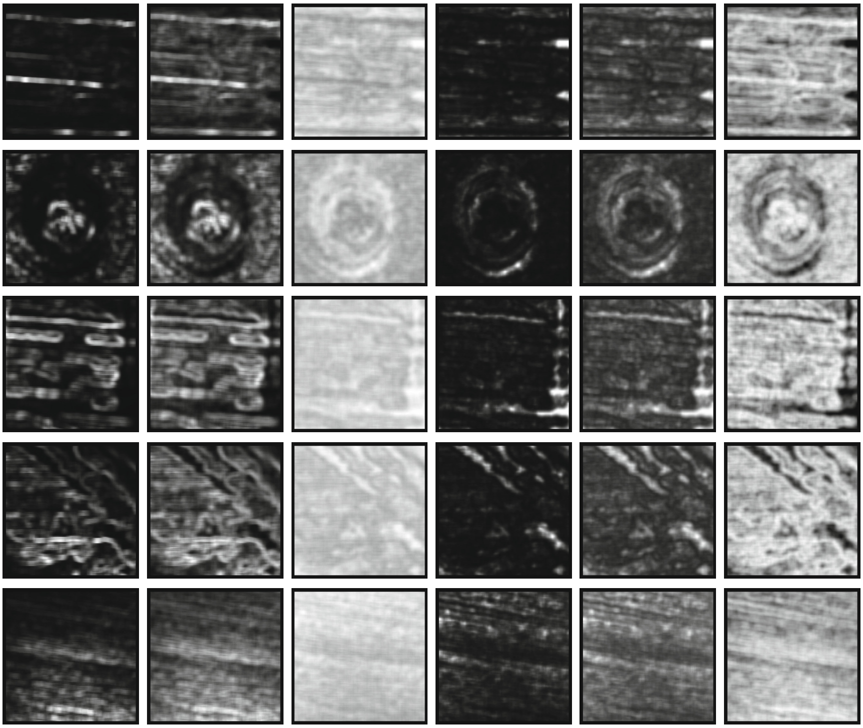


**Fig. 5** Texture statistics of the images bullet striation mark. Each row corresponds to different images with Figure 3. From top to bottom: bullet, firing pin, extractor mark, ejector mark, cartridge. Images from left to right: contrast image, dissimilarity image, homogeneity image, asm image, energy image and entropy image. Window size $w = 21$, $d = 5$.

- Dissimilarity.

$$Dis = \sum_{k=0}^{G} \sum_{l=0}^{G} C_d(k,l)[|k-l|]^2 \tag{3}$$

- Homogeneity.

$$IDM = \sum_{k=0}^{G} \sum_{l=0}^{G} \frac{C_d(k,l)}{1+(k-l)^2} \tag{4}$$

- Angular Second Moment.

$$ASM = \sum_{k=0}^{G} \sum_{l=0}^{G} C_d(k,l)^2 \tag{5}$$

- Energy.

$$Energy = \sqrt{ASM} \tag{6}$$

- Entropy.

$$Entropy = \sum_{k=0}^{G} \sum_{l=0}^{G} C_d(k,l)(-\ln C_d(k,l)) \tag{7}$$

## 3 Experimental Evaluations

The initial experiments were designed to establish the image database. The classes used in the experiments were *fired bullets*, *firing pin*, *extractor marks*, *ejector marks* and *cartridge*. In each classes were two original images. Each original image is split into 5 smaller tiles. In this case the database contains 50 images. A query image is selected from this database - we randomly picked 1 image from each class and used them as queries. The task is to retrieve the images from database that are most similar to the query image. The retrieval rate $Ret_{rate} = \frac{\#correct\ matches}{Number\ of\ tiles}$. 5 best matches are considered and in our experiment all query images were correctly retrieved.

Image retrieval is done in the following steps. First, a test image is given as a query to the image retrieval system. Second, apply a feature extraction algorithm to the test image, and obtain a feature vector. Four texture feature descriptors were used in the experiment: contrast, angular second moments, homogeneity, entropy. The experimental results for images with Figure 3 are listed in Table 1. Third, apply the distance measure used Euclidean distance.

**Table 1** Texture parameters of the images with Figure 3

| | Fig.3a | | Fig.3b | | Fig.3c | | Fig.3d | | Fig.3e | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | $\theta$ | $d = 5$ | $\theta$ | $d = 5$ | $\theta$ | $d = 5$ | $\theta$ | $d = 5$ | $\theta$ | $d = 5$ |
| Contrast | 0 | 565.549 | 0 | 2140.270 | 0 | 329.856 | 0 | 890.629 | 0 | 501.608 |
| | 90 | 1603.831 | 90 | 2880.853 | 90 | 1373.509 | 90 | 1533.689 | 90 | 1082.068 |
| IDM | 0 | 0.149 | 0 | 0.047 | 0 | 0.138 | 0 | 0.110 | 0 | 0.101 |
| | 90 | 0.061 | 90 | 0.051 | 90 | 0.083 | 90 | 0.081 | 90 | 0.055 |
| ASM | 0 | 2.633E-4 | 0 | 8.083E-4 | 0 | 5.477E-4 | 0 | 4.946E-4 | 0 | 2.800E-4 |
| | 90 | 1.159E-4 | 90 | 7.665E-5 | 90 | 3.765E-4 | 90 | 3.884E-4 | 90 | 1.763E-4 |
| Entropy | 0 | 8.929 | 0 | 9.729 | 0 | 8.889 | 0 | 9.114 | 0 | 8.620 |
| | 90 | 9.623 | 90 | 9.847 | 90 | 9.275 | 90 | 9.400 | 90 | 9.042 |

## 4 Conclusion

The paper is devoted to specific topics and problems of firearm forensics image processing and retrieval. A special attention is devoted to comparison of marks on firearm bullets and extract feature vectors that represent striations characteristics. Proposed algorithm of image retrieval use *GLCM* matrices and some texture parameters. Experimental results showed that the proposed method is effective and suitable for forensic image retrieval. The future work focuses on methods for the extraction low and high-level features which might allow for the automatic/computer identification of firearm forensics images.

## References

1. Saferstein, R.: Criminalistics. An introduction to Forensic Science. Printice Hall (2007)
2. Thompson, R.M.: Firearm Identification in Forensic Science Laboratory. NDAA (2000)
3. Choraś, R.S., Andrysiak, T., Choraś, M.: Content Based Image Retrieval Technique. Pattern Analysis and Applications 10, 333–343 (2007)
4. Huang, Z., Leng, J.: A novel binarization algorithm for ballistics imaging systems. In: Proc. 3rd Int. Congress on Image and Signal Processing, CISP 2010, pp. 1287–1291 (2010)
5. Huang, Z., Leng, J.: An online ballistics imaging system for firearm identification. In: Proc. 2nd Int. Conf. on Signal Processing Systems, ICSPS, pp. 68–72 (2010)
6. Li, D.G.: Ballistics projectile image analysis for firearm identification. IEEE Trans. on Image Processing 15, 2857–2865 (2006)
7. Geradts, Z., Zaal, D., Hardy, H., Lelieveld, J., Keereweer, I., Bijhold, J.: Pilot investigation of automatic comparison of striation marks with structured light. In: Proc. SPIE, vol. 4243, pp. 49–56 (2001)
8. Choraś, R.S.: Computer Vision. EXIT, Warszawa (2005) (in Polish)

# An Algorithm for Objects Straightening in M-Mode Echocardiography Images

Adam Piórkowski

Department of Geoinfomatics and Applied Computer Science
AGH University of Science and Technology, Cracow, Poland
pioro@agh.edu.pl

**Summary.** This article presents an approach of M-Mode image analysis. These images are important for the diagnosis of heart valves . This implicates a need for accurate object analysis. A special preprocessing is needed. There is an algorithm for objects straightening proposed in this article. This transformation enables correct classification of objects. Considerations are illustrated by examples.

## 1 Introduction to M-Mode Echocardiography Images

The M-Mode is a 1D variation of B-mode (2D) ultrasound used in echocardiography. It results in displaying a motion of a reflector over distance and time. M-mode echocardiography is helpful in evaluating the morphology of structures, movement and velocity of cardiac valves and walls and timing of cardiac events.

There are several options for the use of M-Mode ultrasound. One of them is to check the operation of valves. M-mode images contain information about emerging, along with blood flow bubbles of gas. These examples assume that the X-axis is the axis of time, and the Y-axis corresponds to the diameter vessel (aorta). Figure 1 presents an example of M-Mode image. In this case, small dots correspond to the fast-moving bubbles, while horizontal stripes - bubbles staying for a long time in the range of ultrasound beam (no flow).

## 2 Problems of M-Mode Image Analysis

Analysis of M-Mode images should give effects in the form of precise count of all objects (bubbles). The result should be in the form of a graph with the number of objects in successive moments of time. It should be noted that the streaks are generated by a single object and we can accept their presence as a point at the start.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 33–38.
springerlink.com 

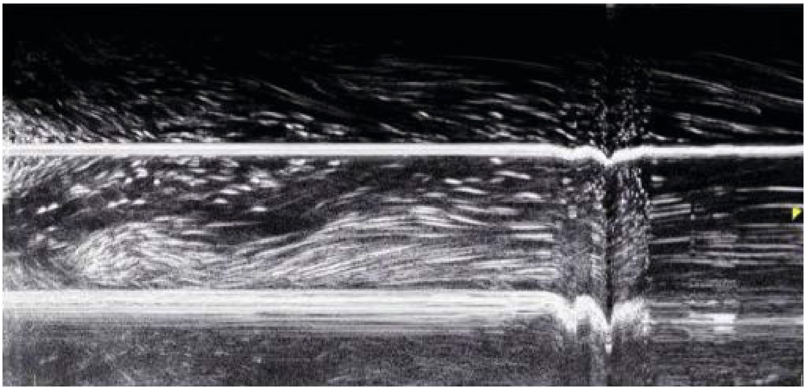


**Fig. 1** An example of a M-Mode image

There are a few approaches to M-Mode image analysis.

The authors describe a method for automatic defining conduit boundaries from a synthetically generated M-Mode image [1]. They use spline curves that are generated from the neighboring conduit boundaries, this technique estimates the LVOT diameter even in a noisy image where conduit boundaries are not clear.

The problem of segmentation in M-Mode images was presented by [2]. The authors propose a new methodology that automatically segments all the walls present in M-mode echocardiography images. The purpose is to determine the ventricular chamber dimensions and wall thickness. The method is based on Amplitude-Modulation Frequency Modulation (AM-FM) demodulation techniques.

Another article [3] describes a method for segmenting arterial vessel wall motion to sub-pixel resolution. The main goal of this article is measuring the spatial offset between all pairs of scans from their crosscorrelation, converting the spatial offsets to relative wall motion through a global optimization, and finally translating from relative to absolute wall motion by interpolation over the M-Mode image. The authors consider mostly the moving borders (wessel wall) than the moving bubbles of gas.

In paper [4] authors present methods for detection of temporal motion velocity and acceleration of omnidirectional M-mode echocardiography. There is obtainted temporal velocity and acceleration of one part of the cardiac structure using one-order and two-order differential of the discrete function.

The most interesting consideration, related to topic of this article is the work [5]. The authors processed a detection of bubble trajectories in M-Mode images. A Kalman filter model is used to estimate the instantaneous values of variables such as the bubble position, velocity and acceleration, in a discrete time process. The solution is achieved by two control parameters: the variance of acceleration and the maneuverability, characterizing the bubble inertia.

### 2.1 Binarization

Binarization is a first preliminary stage (preprocessing). The use of threshold binarization does not give good results because of the varying intensity of the brightness of the bubbles, which is related to the distance from the stream of ultrasound. Tests showed that satisfactory results are achieved by using adaptive binarization algorithm, based on the context of an object (neighborhood) and automatically selecting the threshold at the half of the brightness range of this context. The shape of the context area was set to a square of 11x11 pixels or to horizontal line of 20 pixels.

### 2.2 A Problem of Object Identification

The problem of identifying the objects is associated with artifacts that occurs during the test. The figure 2 shows a part of the study of selected sites A and B. The A circle contains two streaks of two objects. The brightness of these objects makes the binarization and indexation may be join these two object into one. The next problem is the issue of combining objects. Case B shows three objects giving the streaks, which are not continuous. Their separation is a mistake because it creates additional objects that do not occur during the flow.

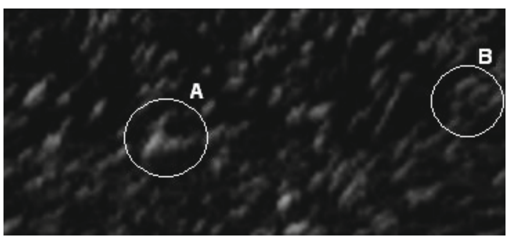



**Fig. 2** Artifacts in M-Mode images

## 3 A Proposal of an Algorithm for Objects Straightening

The presented problem of improper joining or separating objects can be resolved using the knowledge about the flow direction. Then every object that creates a streak parallel to the direction of flow can be followed in order to connect objects, and objects on a line perpendicular to the direction of flow should be separated. This idea seems to be correct, but there is a problem of variable directions of flow in the aorta. Therefore it is reasonable to straighten the flow image, and then apply filters to join probably continuous objects in the horizontal and to separate distinct objects in the vertical (directional filters).

### 3.1 An Algorithm for Objects Straightening

Each column contains a full section of dishes in a given time. The aorta is waving, so in subsequent moments of time the position is changing. From this it follows that the columns are shifted relative to each other. The idea behind the algorithm is to find the offset and compensate for it to straighten the image.

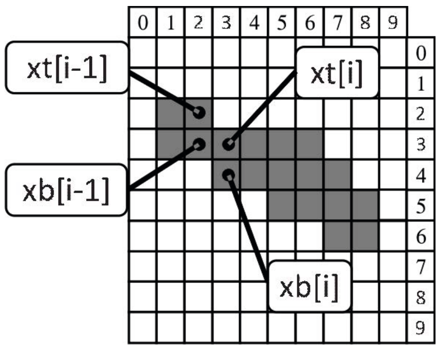



**Fig. 3** An idea of algorithm for straightening M-Mode images

The figure 3 shows a sample object in the image. Suppose we examine the column No. 3 (i). Then it is necessary to compare how the object slices moved between the time corresponding to the column 2 (i-1) and just next to the column 3 (i). The top point of the object in $i$ column is described by row $xt_i$, and the bottom - by $xb_i$. The values for the previous column $(i-1)$ should be calculated analogously.

Center points of objects for columns $xm_i$ and $xm_{i-1}$ are determined by the formulas (1) and (2):

$$xm_i = xt_i + \frac{xb_i - xt_i}{2}. \tag{1}$$

$$xm_{i-1} = xt_{i-1} + \frac{xb_{i-1} - xt_{i-1}}{2}. \tag{2}$$

Column offset for a single object ($objectshift_i$) is the difference of these center points according to the formulas (3) and (4):

$$objectshift_i = xm_i - xm_{i-1}. \tag{3}$$

$$objectshift_i = xt_i + \frac{xb_i - xt_i}{2} - xt_{i-1} - \frac{xb_{i-1} - xt_{i-1}}{2}. \tag{4}$$

The value of column offset for the object on Fig. 2 is:

$$objectshift_3 = xt_3 + \frac{xb_3 - xt_3}{2} - xt_2 - \frac{xb_2 - xt_2}{2} = 3 + \frac{4-3}{2} - 2 - \frac{3-2}{2} = 1. \quad (5)$$

To calculate the average column offset ($columnshift_i$) the average of offsets should be determined for all the objects appearing in this column:

$$columnshift_i = \sum_{k=1}^{m} \frac{objectshift_i}{m}. \quad (6)$$

The next step is to define the absolute shift of columns ($absshift_i$), which is the sum of the average column offsets for all preceding columns from the beginning of the measurement:

$$absshift_i = \sum_{k=1}^{i} columnshift_i. \quad (7)$$

The final step is to create a new image which will take account of shift columns. The values of points $pn[i, j]$ in the new image from the original image $p[i, j]$ are given by the formula:

$$pn[i, j] = p[i, j - absshift_i]. \quad (8)$$

It should be noted that the offset is the floating point number, so the original image needs to be created by interpolating shifts:

$$\begin{aligned} pn[i, j] = {} & (1 - m(absshift_i)) * p[i, j - absshift_i] + \\ & (m(absshift_i)) * p[i, j - absshift_i + 1]. \end{aligned} \quad (9)$$

### 3.2 Using the Proposed Algorithm

The figure 4 presents the indexed original M-Mode image, and figure 5 is the result of the algorithm for this image. It can be noticed that the objects obtained after processing mostly keep the horizontal position, so the proposed approach achieved the objective.

**Fig. 4** An example of indexed M-Mode image

**Fig. 5** Indexed M-Mode image after straightening

## 4 Conclusions

The presented algorithm fairly well straightened image consisting of the indexed objects. Most of the objects adopted a horizontal position. Such preparation allows you to perform further analysis. Further work will be focused on the development of this method and analysis for clinical trials of M-Mode. An interesting approach can be realized using swarm dynamic analysis [6].

**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. Kim, B., Stamos, T.D., Soble, J.S., Roberge, J.: Automated Conduit Detection Method from a Synthetic M-mode Using Spline Curves. IEEE Computers in Cardiology 21(27), 21–24 (2000)
2. Rodriguez, P.V., Pattichis, M.S., Goens, M.B.: M-mode Echocardiography Image and Video Segmentation based on AM-FM Demodulation Techniques. In: Proceedings of the 2Sb Annual International Conference of the IEEE EMBS Cancun, Mexico, September 17-21 (2003)
3. Fancourt, C., Azer, K., Ramcharan, S.L., Bunzel, M., Cambell, B.R., Sachs, J.R., Walker, M.: Segmentation of arterial vessel wall motion to sub-pixel resolution using M-mode ultrasound. IEEE Eng. Med. Biol. Soc. (1), 3138–3141 (2008)
4. Lin, Q., Lin, Y., Wang, H.Y., McAllister, H.G., Zheng, H.: Detection of Temporal Motion Velocity and Acceleration of Omnidirectional M-Mode Echocardiography. Computers in Cardiology (2006)
5. Balocco, S., Basset, O., Guidi, F., Tortoli, P., Cachard, C.: Detection of microbubble trajectories on M-mode images using Kalman filtering. In: IEEE ICASSP, Toulouse, pp. 569–572 (2006)
6. Gajer, M.: Visualization of particle swarm dynamics with the use of Virtual Reality Modeling Language. Electrical Review 87(11), 20–24 (2011)

# Image Registration Tuning for DCE-MRI Breast Imaging

Karol Kuczyński[1], Maciej Siczek[2], and Rafał Stegierski[1]

[1] Maria Curie-Sklodowska University, Pl. M. Curie-Sklodowskiej 1, 20-031 Lublin, Poland
karol.kuczynski@umcs.pl
[2] Hospital of Ministry of Interior and Administration, ul. Grenadierow 3, 20-331 Lublin, Poland

**Summary.** DCE-MRI is a novel medical imaging technique for breast cancer diagnostics. Unintentional patient's movements during imaging session result with misalignments between consecutive image series. Their analysis is then problematic. The problem can be solved by application of image registration procedure. A system for DCE-MRI breast image registration, using B-spline transformation has been created. The paper presents work on its testing and tuning, to improve both performance and accuracy.

## 1 Introduction

Breast cancer mortality has been seriously reduced in the recent years, but still it is a vital social problem. According to the World Health Organization's reports, *breast cancer is the leading cancer killer among women aged 20-59 years in high-income countries* [1]. Its diagnostics is currently based mainly on mammography and ultrasonography. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a novel technique that offers superior sensitivity [2, 3]. It provides information on tissue properties, including tumour status that is available neither with mammography, USG nor the regular MRI. A few series of MR (magnetic resonance) images of the same body region are rapidly acquired before, during and after injection of paramagnetic contrast agent (Gd-DTPA). Propagation of the contrast agent modifies MR signal over time that is then analysed.

A patient should remain statically during the whole imaging session, but it is not always possible, due to various factors. Unintentional movements result with misalignment of consecutive image sequences. It makes further analysis difficult or impossible. Repetition of the whole examination is problematic because it is time-consuming, relatively complicated and expensive. Application of image registration (fusion) techniques to properly align corrupted images seems to be a natural solution.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 39–46.
springerlink.com 

General approach to the problem is quite obvious. Adequate algorithms are known. However, design of a registration system that could be used in a routine manner in a hospital is a much more complex problem than just registration of carefully selected image datasets. The proposed registration framework has been already created and presented [4, 5]. The purpose of the work presented in this paper is to tune its parameters and apply additional mechanisms to achieve better performance and accuracy. The expected goal is to make a system that provides satisfactory results within acceptable time, using commonly available computer hardware, for image datasets acquired in a routine manner.

## 2 Materials and Methods

### 2.1 Image Datasets

The image collection currently consists of images of breast from more than 100 dynamic MR sessions. Most of them are perfectly aligned thanks to a special patient positioning system. However, up to a few percent of the images may be problematic due to patients' movements. Each session consists of six consecutive image sequences showing the same body fragment. Each sequence consists of about sixty T1 FATSAT axial 512 × 512 pixels slices (fig. 1.).

The first sequence is treated as a fixed image for the registration process. The other five sequences are moving images, to be registered with the first one. Thus, the task consists of five separate registration subtasks. The resulting images are then composed into a new DICOM dataset, preserving

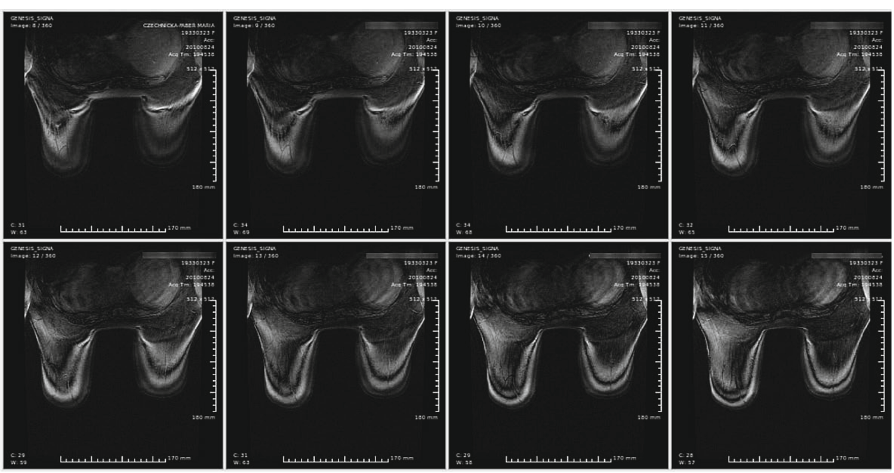


**Fig. 1** Fragment of the 1st sequence (out of six) of a dataset

timing information and other study details that are used by DCE-MRI analysis software.

### 2.2 Algorithms

A great variety of registration algorithms in currently known [7, 8, 9]. However, election of registration framework main components for the problem was quite straightforward. Three-dimensional registration needs to be employed. Subtle local deformations are expected, rather than large rotations, translations, scaling or shearing distortions. It is already known that deformable transform using a B-spline [11] representation is appropriate for breast image registration [12, 13].

In a multi-modal registration (where correspondence between greylevels in both images is not evident) mutual information is commonly employed as similarity measure. Although some pixel intensities change during a DCE-MRI session, it has been shown [5] that a simple mean squares metric [14] (much less computationally expensive than mutual information) still can be used:

$$MS(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2 \qquad (1)$$

The optimisation process is done with LBFGSB [15] algorithm. In the last step, a moving image is transformed using the final transformation parameters, resampled using linear interpolation and saved in the fixed image space.

Image registration is a time-consuming procedure. Considering that most of currently available PCs have a few processor cores and the whole registration can be divided into 5 subtasks, it is reasonable to run them in parallel threads. Time gain resulting from the parallelisation (implemented with Boost library) is very significant and has been discussed in [5].

Multi-resolution is a widely used technique to improve both speed and quality of a registration process. Registration is at first performed with smaller (scaled down) images. When the registration criterion is reached, the calculated transformation parameters are used as a starting point for the next step, with larger images. This procedure is usually repeated until the final registration performed with full-scale images.

In the presented framework with B-spline transformation, there are at least two possible ways of multi-resolution implementation. Either several resolutions of the image itself or several resolutions of B-spline grid can be used. Both approaches have been tested. The other possibility is a combination of the two methods mentioned above.

## 3 Results

Fig. 2 presents pixel-wise squared difference between corresponding images in one of the most problematic sessions, before and after the registration

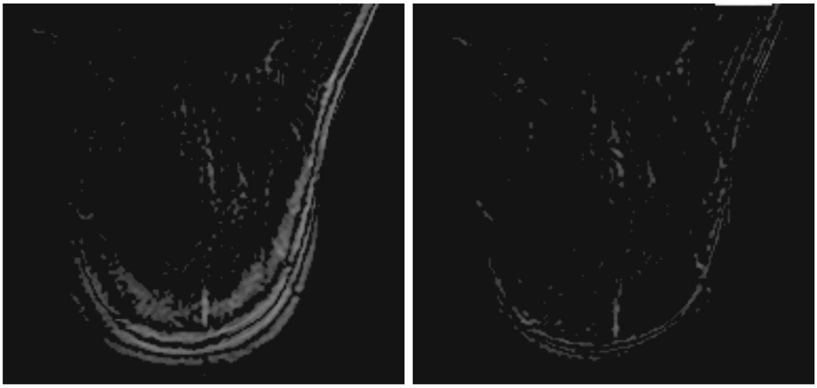


**Fig. 2** Pixel-wise squared difference between corresponding slices in the 1st and 3rd sequence, in the original dataset (left) and after the registration (right)

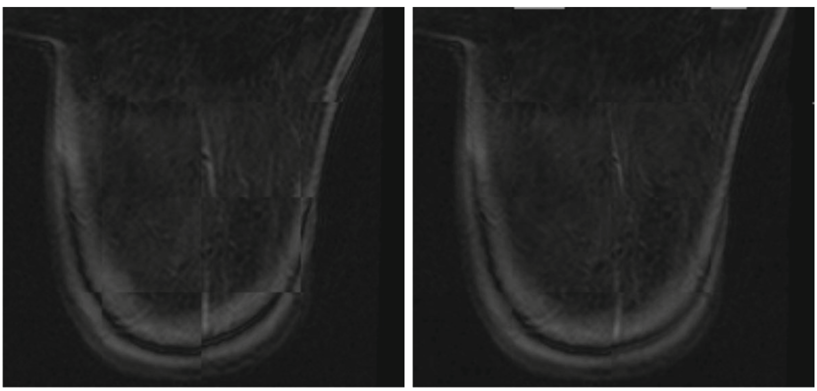


**Fig. 3** A checkerboard test for corresponding slices in the 1st and the 3rd sequence, for original slices and after the registration

process. The displacement is significant and also well visible in the checkerboard pattern created from the images (fig. 3).

The registration was performed on the region of interest only, using B-splines of order 3 and 10-nodes grid size in each dimension. The optimizer's stop condition was gradient tolerance equal to 0.05. The five registration subtasks were performed in parallel threads (using Boost library), as described in [5].

The graph in fig. 4 presents the optimisation process, with and without multi-resolution approach. In the first (out of two) multi-resolution level, image size was reduced by 50% in each direction. The second level was

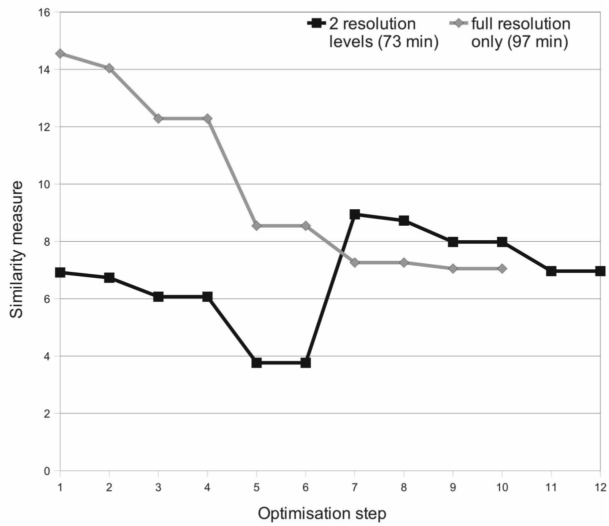



**Fig. 4** Optimisation process, with multi-resolution approach (black line, steps 1–6 are performed with reduced resolution, 7–12 with full resolution) and without it (grey line), for one image pair

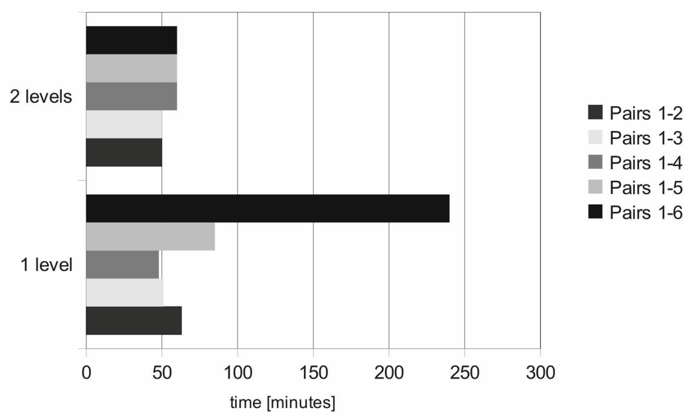



**Fig. 5** Optimisation time, with multi-resolution approach (2 levels) and without it (1 level), Intel®Core™i7, 3.2 GHz computer; all image pairs registered in parallel

performed using full resolution images. Using more than two multi-resolution levels is useless, because registration of low-resolution images quickly converges to identity transformation. Application of multi-resolution approach usually results with shorter processing time (in the presented case, the time gain is uncommonly big: 73 minutes instead of 97 minutes, Intel®Core™i5 M520, 2.4 GHz computer). Similar results can be obtained by changing the B-spline grid size (7 nodes in the first level, 10 nodes in the second level), instead of image resolution itself.

Fig. 5 illustrates another problem. All image pairs were registered in parallel. Registration of the 1st ant 6th pair was problematic due to presence of local minima. The registration was successful but the whole task took much time. The problem did not take place when multi-resolution was used.

## 4 Discussion

Experiments similar to those presented in the above section have been performed on most images from the collection. Registration progress and time vary from image to image and cannot be accurately predicted. However, usually image displacements are much smaller (and thus easier to correct) than the presented in fig. 2–3.

Multi-resolution application is expected to reduce computing time and to improve accuracy. In most cases the results were compliant with the expectations (the time was reduced by a few percent, a slightly deeper minimum of the registration criterion was found). For a few percent of the datasets, no positive influence has been observed. Generally, multi-resolution approach makes the registration process more stable and predictable, and reduces risk of falling into local minimum (fig. 5).

In most of the multi-resolution experiments, resolution of images was changed, but changing of B-spline grid nodes has been also tested. This issue still needs to be tested more thoroughly, but it seems that after parameters' tuning, very similar results may be obtained with both approaches.

Registration accuracy can be modified by setting the B-spline grid size and stop criterion. Better accuracy costs more processing time. For $190 \times 180 \times 60$ pixels region of interest, reasonable grid size in at least 10 B-spline grid nodes in each direction (with gradient tolerance of 0.05), considering that registration time on a standard, modern hardware PC (Intel®Core™i7, 3.2 GHz computer) for the whole dataset is about 1 hour. For 15-nodes grid (and gradient tolerance of 0.02) it is more than 10 hours.

## 5 Conclusion

From a practical point of view, the result of parallelisation of the registration process was the most spectacular. The computing time was reduced by more than 50% (depending on hardware used), for all of the tested datasets [4, 5]. Performing the registration on the region of interest only (instead of the whole image) also greatly reduces computing time or makes it possible to increase accuracy (use more B-spline nodes), but the region is to be manually selected.

The effect of multi-resolution approach is more difficult to judge, because it is different for various datasets. It was expected to reduce computing time or to improve accuracy. In most cases the results were compliant with the expectations, but the gain was rather subtle. It is more important to notice that application of multi-resolution greatly reduces risk of problems related to local minima. Their occurrence drastically increases registration time or prevents successful solution.

**Acknowledgement.** MR image data used in this study has been provided by the Hospital of Ministry of Interior and Administration in Lublin (Poland).

## References

1. Women's health, World Health Organization, Fact sheet No. 334 (November 2009), http://www.who.int/mediacentre/factsheets/fs334/en/
2. Ojeda-Fournier, H., Comstock, C.E.: MRI for breast cancer: Current indications. Indian J. Radiol. Imaging 9, 161–169 (2009)
3. Li, X., Dawant, B.M., Welch, E.B., Chakravarthy, A.B., Freehardt, D., Mayer, I., Kelley, M., Meszoely, I., Gore, J.C., Yankeelov, T.E.: A Nonrigid Registration Algorithm for Longitudinal Breast MR Images and the Analysis of Breast Tumor Response. Magn. Reson. Imaging 27(9), 1258–1270 (2009)
4. Kuczynski, K., Siczek, M., Stegierski, R.: Application of Image Registration Techniques in Magnetic Resonance Imaging of Breast. Journal of Medical Informatics & Technologies 17 (2011)
5. Kuczyński, K., Siczek, M., Stegierski, R.: DCE-MRI Breast Image Registration for Tumour Diagnostics. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 3. AISC, vol. 102, pp. 323–329. Springer, Heidelberg (2011)
6. Viola, P., Wells, W.M.: Alignment by Maximization of Mutual Information. International Journal of Computer Vision 24(2), 137–154 (1997)
7. Zitova, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing 21, 977–1000 (2003)
8. Antoine Maintz, J.B., Viergever, M.A.: A Survey of Medical Image Registration. Medical Image Analysis 2(1), 1–37 (1998)
9. Goshtasby, A.: 2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications. Wiley Press (2005)
10. Kuczynski, K., Mikolajczak, P.: Information theory based medical image processing. Opto-Electronics Review 11(3), 253–259 (2003)
11. Unser, M.: Splines: A Perfect Fit for Signal and Image Processing. IEEE Signal Processing Magazine 16(6), 22–38 (1999)
12. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast mr images. IEEE Transaction on Medical Imaging 18(8), 712–721 (1999)

13. Mattes, D., Haynor, D.R., Vesselle, H., Lewellen, T.K., Eubank, W.: PET-CT image registration in the chest using free-form deformations. IEEE Trans. on Medical Imaging 22(1), 120–128 (2003)
14. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide, Updated for ITK version 2.4, 2nd edn. (2005), http://www.itk.org
15. Byrd, R.H., Zhu, C., Nocedal, J.: L-bfgs-b: Algorithm778: L-bfgs-b, fortran routines for large scale bound constrained optimization. ACM Transactions on Mathematical Software 23(4), 550–560 (1997)

# Variogram Based Estimator of Fractal Dimension for the Analysis of Cell Nuclei from the Papanicolaou Smears

Dorota Oszutowska–Mazurek[1], Przemysław Mazurek[2], Kinga Sycz[3], and Grażyna Waker–Wójciuk[3]

[1] Gryfice Hospital Medicam, Department of Pathomorphology, Niechorska 27 Str., 72300 Gryfice, Poland
adorotta@op.pl

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

[3] Independent Public Voivodeship United Hospital, Department of Patomorphology, Arkońska 4 Str., 71455 Szczecin, Poland
grazynka@blue.net.pl

**Summary.** The aim of cervical screening is detection of potentially pre–cancerous and cancerous changes in the endocervical canal of female reproductive system. There are many Fractal Dimension (FD) that are important for the computer assisted analysis of the texture of cell nuclei. The best method is the algorithm that process all pixels of the cell nuclei, not limited by the constraints. Variogram algorithm is proposed for cell nuclei classification (correct and atypical cell classes). Obtained results shows that variogram and cell nuclei size are important for the automatic classification of cell nuclei.

## 1 Introduction

The aim of cervical screening is detection of potentially pre–cancerous and cancerous changes in the endocervical canal of female reproductive system. The introduction of screening test decreased the incidence and mortality of cervical cancer in some populations, where women take part in the screening program regularly. Cytoscreeners and pathomorphologists engaged in this process look for abnormalities [4].

Specific stain – Papanicolaou method is provided in this procedure and such smears are observed (Fig. 1) under the microscope mainly by cytoscreeners and pathomorphologists, who confirm uncertain and positive cases like

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 47–54.
springerlink.com 

precancerous conditions(cervical intraepithelial neoplasia) and cancer. Certain positive cases are confirmed by histopathological examinations.

The analysis requires specialists with adequate experience after training in Papanicolaou smears screening, because biological objects like cells observed under the microscope are variable and complex. The variability is caused by many factors like the patient age, lifestyle, presence of infections, etc. The Papanicolaou smears [5, 6] are evaluated according to Bethesda system for reporting cervical cytology, revisited in 2001 [15, 2]. The specialists assess not only cancerous conditions, but other diagnostic phenomena. The errors in analysis may be connected even with the risk of the patient life. During Papanicolaou smears screening different properties of cells and especially cell nuclei are considered by the cytoscreener.

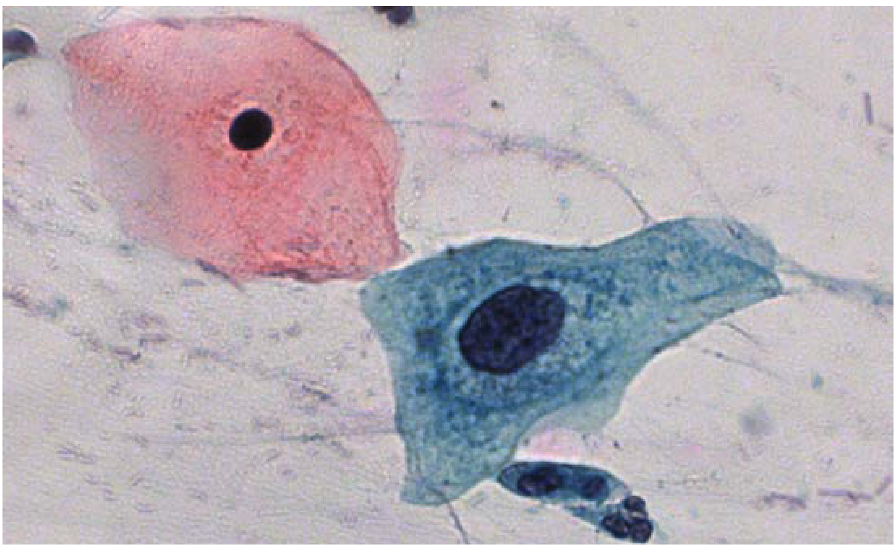


**Fig. 1** Example of Papanicolaou smear

The few main features of cells areobserved – the nuclear–cytoplasmic ratio (N/C ratio), connected with the size of cell nucleus, the texture and contour of the nucleus. Benign cells have regularly shaped nuclei with smooth contour and the N/C ratio is not increased. The opposite features are observed in case of neoplastic or cancer cells: the nuclei are irregular with grainy texture, pleomorphic, hyperchromatic, rough contour and the N/C ratio is high.

## 2 Fractal Dimension Estimators for Optical Density

There are many Fractal Dimension (FD) estimators [8, 7, 13] that are used for analysis of the texture of cell nuclei. The cell nuclei have different sizes. An examples of cell nuclei are shown in Fig.2

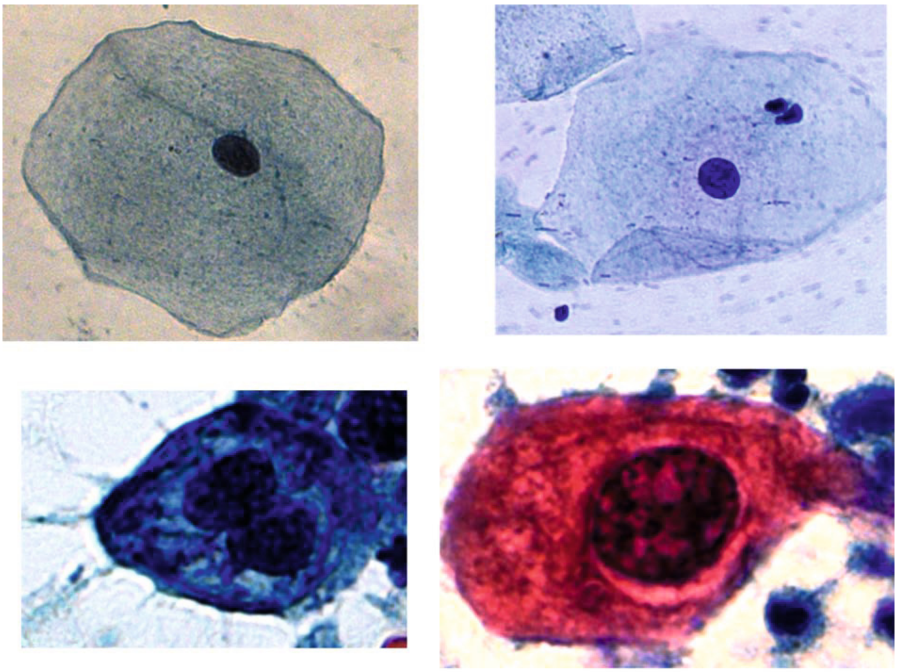


**Fig. 2** Example shapes of cells. Correct cell nuclei (top row). Atypical cells nuclei (bottom row). Scale is not preserved.

Analysis of the shape of cell nuclei contour is possible using different shape descriptors. The FD estimators for shape are considered in [10, 11], and the perimeter changes due to scale are used. Another possibility is the FD estimation of the field of cell nuclei due to scale and the box counting technique is used.

Such techniques are not related to the texture. Granularity of the cell is considered also and it is important factor for cytoscreener. The FD estimation for optical density is necessary.

The TPM (Triangular Prism Method) is the main technique for the estimation of the optical density [3, 17, 18]. This algorithm is applied for the estimation of the landscape parameters, but this method is more general and not related to the unique application. The TPM considers the height of the landscape at regular grid points (rectangular or triangular based). In the case of the microscopic image there are not measured a heights (thickness) of the cell nucleus. Conventional microscopy is based on the transmissive measurements. The dark pixel is related to the very dense area, and white pixels to the translucent area. There is relation between thickness of cell nuclei and optical density for particular pixel, but unknown. The conversion between optical density and the thickness is not necessary for computer assisted diagnosis of Papanicolaou smears [19]. There are optical effects like optical interference, diffraction, and many other that are visible, and they change pixel values. This is another reason why not a thickness, but the optical density is used.

The TPM algorithm is simple for computations, but the most important limitations are constraints related to the area of analysis. The TPM algorithm processes square area ($N$x$N$) that have side $N$, defined by the following formula:

$$N = 2^s + 1 \tag{1}$$

The area of cell nucleus is not squared so not all pixels of cell nuclei are covered by the TPM algorithm. Two main problems are important. Omitting of area outside of TPM analysis square influences the quality of results. The reduced numbers of pixels, that are used for FD estimation increase the variance of FD estimator (local FD and global FD [14]). The grid of the TPM algorithm should be fitted to the largest possible square inside the cell nuclei. This is not possible without interpolation, and image rescalling also influences the results.

The best method is the algorithm that process all pixels of the cell nuclei, not limited by the constraints. Computation of the local FD is required, because the textures may have different local FD's depending on the scale.

## 3 Variogram Based Estimator of Fractal Dimension

The variogram is calculated, using the following formula:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[z(x_i) - z(x_i + h)\right]^2 \tag{2}$$

The $h$ is the Euclidean distance between two samples ($z(x_i)$ and $z(x_i + h)$). The $N(h)$ is the number of the samples that have the same distance $h$. The variogram is a case of the histogram and has similar problems for precise data sets. There are large number of $h$ distances that are unique, so approximation of the envelope is necessary. In this paper the scale value equal to 1 corresponds to the Euclidean distance equal to $h = 3$. Such approximation is sufficient.

The formula (2) is independent on the data set, so even non–uniform sampling cases are supported. The area of interest (cell nucleus) could be of any shape and all pixels of cell nuclei are processed. This is very important advantage over other techniques.

## 4 Experimental Results

The database of cell nuclei of Papanicolaou smears consists images of cell nuclei and masks for particular cells. The microscope with lenses 40 (magnification 400x) is used. Papanicolaou smear image is acquired using AxioCamMRc5 camera that supports 2584x1936 resolution. Color RGB images are obtained, but the only green channel is used. The green channel consists the most important information about texture and has higher accuracy, because the camera has Bayer sensor [12].

In Fig. 3 and Fig. 4 are shown variograms for cell nuclei. It is well visible that variograms are different, but some shapes are similar. The size of cell nucleus is also important factor that must be used in the analysis. Instead of FD formula:

$$FD = 2 - \frac{\theta}{2} \tag{3}$$

where $\theta$ is the slope of the regression line [16] between $\log^*(h)$ and $\log(h)$, the differences between variograms values for specific $h$ distances are used (Fig. 5).

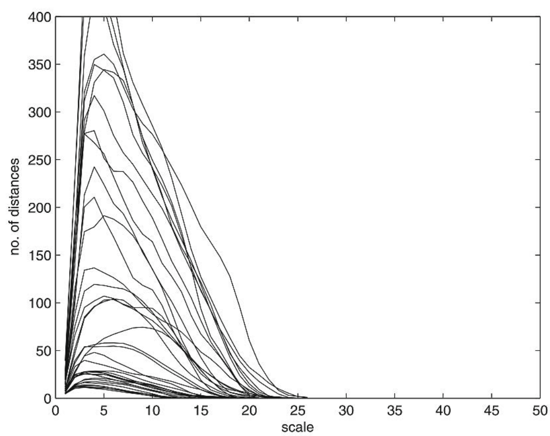



**Fig. 3** Variograms of correct cell nuclei

The classification is possible using almost linear classified. Only one atypical cell nuclei is located between correct cells nuclei. There are two populations of atypical cells. The first one is characterized by the large field area. The second one has size similar to the correct one.

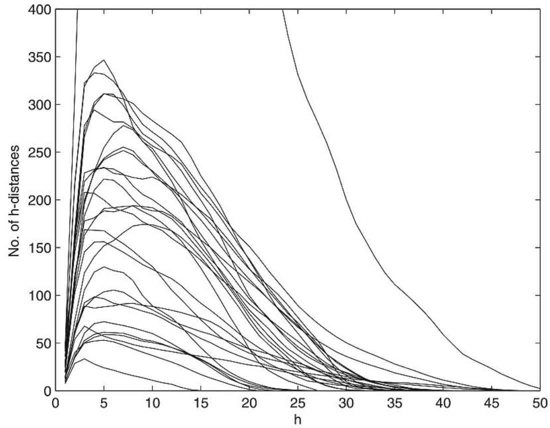



**Fig. 4** Variograms of atypical cell nuclei

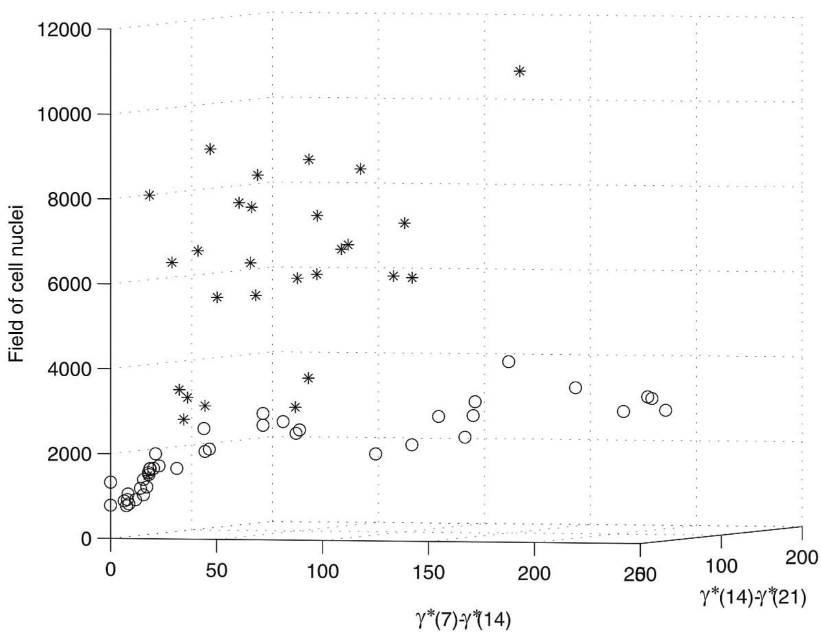



**Fig. 5** Classification space for correct and atypical cell nuclei. Scales 7 ($h = 21$),14 ($h = 42$), 21 ($h = 63$). Correct cells nuclei are marked as circles. Atypical cells nuclei are marked as stars.

## 5 Discussion and Conclusions

Automatic screening [9, 1] of the cell nuclei of Papanicolaou smears would be possible. The obtained results (detected atypical) cells must be verified by the cytoscreener and pathomorphologist additionally. Such system may work as a preclassification system. The final classification is provided by human experts. Detection of the atypical cells nuclei allows reduction of the miss ratio and could improve overall process.

Application of the variogram as tool for analysis of cells is possible and necessary. All pixels are used for estimation of FD or using simpler difference technique. The selection of the scales of variogram is not sensitive, and three of them are arbitrary selected. Selection of the scales is based on the variograms (Fig. 3 and Fig. 4). The minimal value (7) is similar to the peak of variogram. The largest value (21) is selected as mean of the largest $h$ distance of correct cell nuclei.

Variogram shapes are similar, but there are some additional differences, what is advantage of the variogram technique over other techniques. It is interesting question about possibility of the detection of other features related to the cell nuclei, smear preparation and others.

**Acknowledgement.** This work is supported by the MNiSW grant N514 004 32/0434 (Poland). 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. Blackledge, J., Dubovitskiy, D.: An Optical Machine Vision System for Applications in Cytopathology. ISAST Transactions on Computers and Intelligent Systems 2(1), 95–109 (2010)
2. Chosia, M., Domagała, W.: Cytologia szyjki macicy. Fundacja Pro Pharmacia Futura (2010) (book in Polish)
3. Clarke, K.C.: Computation of the Fractal Dimension of Topographic Surfaces using the Triangular Prism Surface Area Method. Computer and Geosciences 12(5), 713–722 (1986)
4. Dey, P.: Fractal geometry: Basic principles and applications in pathology. Anal. Quant. Cytol. Histol. 27(5), 284–290 (2005) PMID: 16447821
5. Hoda, R.S., Hoda, S.A.: Fundamentals of Pap Test Cytology. Humana Press (2007)
6. IARC, Cytopathology of the uterine cervix – digital atlas. International Agency for Research on Cancer, http://screening.iarc.fr/atlascyto.php
7. Kaye, B.H.: A Random Walk Through Fractal Dimensions. VCH (1994)
8. Mandelbrot, B.B.: The Fractal Geometry of the Nature. W. H. Freeman and Company (1983)
9. McKenna, S.J.: Automated analysis of Papanicolaou smears. PhD Thesis, University of Dundee (1994)

10. Oszutowska–Mazurek, D., Mazurek, P.: Fractal–based generator for microscopic cell and cell nucleus contour synthesis. Metody Informatyki Stosowanej 27(2), 105–112 (2011)
11. Oszutowska–Mazurek, D., Waker–Wójciuk, G., Mazurek, P.: Fractal analysis limitations in digital analysis of Papanicolaou cytological images. Measurement Automation and Monitoring 58(1), 52–54 (2012)
12. Oszutowska-Mazurek, D., Mazurek, P., Sycz, K., Waker-Wójciuk, G.: 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)
13. Peitgen, H.O., Jurgens, H., Saupe, D.: Fractal for the Classroms, Part 1: Introduction to Fractals and Chaos. Springer (1992)
14. Seuront, L.: Fractals and Multifractals in Ecology and Acquatic Science. CRC Press (2010)
15. Solomon, D., Nayar, R.: The Bethesda System for Reporting Cervical Cytology. Springer (2004)
16. Steven, I.: Linear Richardson Plots from Non–Fractal Data Sets. Dutch Mathematical Geology 25(6), 737–751 (1993), doi:10.1007/BF00893176
17. Sun, W., Xu, G., Gong, P., Liang, S.: Fractal analysis of remotely sensed images: A review of methods and applications. International Journal of Remote Sensing 27(22), 4963–4990 (2006), doi:10.1080/01431160600676695
18. Zhou, G., Lam, N.: A comparison of fractal dimension estimators based on multiple surface generation algorithms. Computers & Geosciences 31, 1260–1269 (2005), doi:10.1016/j.cageo.2005.03.016
19. Zieliński, K.W., Strzelecki, M.: Komputerowa analiza obrazu biomedycznego. Wstęp do morfometrii i patrologii ilościowej. Wydawnictwo Naukowe PWN (2002)

# On the Dimensionality of PCA Method and Color Space in Face Recognition

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

**Summary.** The paper presents a problem of color images recognition in the aspect of dimensionality reduction performed by means of different Principal Component Analysis variants. The aim of the experiments was to check the applicability of one- and two-dimensional PCA in color image classification together with an analysis of an employed color-space. Since, most of the works in this area are focused on grayscale images, in this paper we investigate several full-color representations, as it improves the overall recognition rate. As a comparison, typical approaches involving one-dimensional PCA and two-dimensional PCA on RGB, HSV, YIQ and YCbCr representations of images is provided. The paper describes theoretical fundamentals of the algorithm and implementation for these variants of PCA. Furthermore, the impact of the number of principal components on the recognition accuracy is investigated. The usefulness of PCA on color images is investigated on typical benchmark databases containing facial portraits.

## 1 Introduction

The problem of dimensionality reduction in the tasks of image processing and recognition has been present in the scientific literature for many years. Subspace methods, which aim to reduce the dimension of the data while retaining the statistical separation property between distinct classes, have been a natural choice for these tasks [1]. Generally, all subspace methods require input data to be organized in a specific form, which leads to the different strategies of calculations, which finally gives slightly different results. By far, various subspace methods have been proposed and applied to image recognition. One of the most often employed methods is Principal Component Analysis (PCA) also known as Hoteling Transform (HT) or Karhunen-Loeve Transform (KLT), depending of the realization and application. On the other hand, color images of real scenes, are generally high dimensional and their within-class variation is much larger than the between-class variation, which may cause serious performance degradation of classical subspace methods [2].

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 55–63.
springerlink.com 

Principal component analysis has in practice been used to reduce the dimensionality of different data and to transform interdependent coordinates into significant and independent ones [3, 4, 5, 6, 7, 8, 9]. The review of the literature shows that the authors do not deal with a problem of organizing input data for PCA method when such input data are given as high-dimensional structures (more than two-dimensional matrices) and do not investigate the influence of chosen method on the results of their algorithms. While there is an interesting direction of research related to tensor representation of images and the usage of so called High-Order Singular Value Decomposition to preform the projection into the eigenvectors space [10, 11], it is computationally expensive and complicated in the interpretation, thus making a practical implementation unattractive.

The main motivation of this paper is to investigate the influence of dimensionality of traditional PCA method understood as the organization of input data (vectors, matrices) derived from input structures ("three-dimensional" images) on the classification results. We investigate classical one-dimensional PCA (*Eigenfaces* described in [4]) called 1DPCA and simplified two-dimensional PCA (introduced in [12]) called 2DPCA. Although, the simplified two-dimensional PCA is not a "true" two-dimensional transformation (as it was shown in [8]) it is sufficient for comparison purposes. The experiments were conducted on several different color representations (since this problem is often omitted by most of researchers) in order to find the most optimal one.

## 2 Processing Framework

### 2.1 Initial Representation

In many real-life applications we have to deal with color images, since only such images carry whole information about depicted objects. However, most of the classical pattern recognition algorithms are oriented at single-channel images, i.e. intensity representation. Often, the images are transformed from full-color representation (e.g. RGB, HSV) into single-channel form by means of channel separation, then only one channel is chosen for further processing. Sometimes, special transformations procedures are used in order to select most optimal channel to the further processing [13].

In the beginning, input dataset, used to build a database, is divided into two independent parts: training and testing one. The original dimensions of input images are reduced by down-sampling in order to make the computations faster ($r$ is the down-sampling ratio). This stage is introduced to both 1DPCA and 2DPCA. If we define the input image dimensions (after down-sampling) as $M \times N \times D$, then the final reduced representation size is defined as a vector of $m \times 1$ or $m \times D$ elements (of course $m << MN$). In the further parts of the paper we show the simplified processing stages for both methods.

## 2.2 One-Dimensional PCA

### Training

The traditional PCA-based recognition is performed on vector representation of images, which are created using concatenation of rows or columns). In case of grayscale single-channel images this vector's size is equal to $1 \times MN$, while in case of typical color images, no matter which color space we use, the size of the vector is equal to $1 \times MND$. In every color representation investigated here, the parameter $D$ is equal to 3.

First, all image vectors are stored in a common matrix. Then we use the classical PCA algorithm, which performs the following steps:

1. The whole database is employed at the stage of mean vector calculation, followed by normalization (through centering).
2. A covariance matrix of size $MND \times MND$ is calculated for training part of the dataset only, and decomposed into eigenvalues and eigenvectors.
3. According to the typical energetic approach, the $m$ most important eigenvectors are selected to form the transformation matrix of size $MND \times m$.
4. Finally the training images are reduced to the $m \times 1$ size, each.

### Testing

The testing procedure involves reduction of dimensionality of an unknown image and calculating a distance to the nearest object in the training database (classification):

1. The dimensions of test image are reduced by down-sampling according to the parameter $r$.
2. Test image is vectorized and the mean is removed.
3. The test image is projected onto new reduced feature space, and the distances to all images in the database are calculated.
4. The smallest distance, meaning the highest similarity is chosen, thus, the class of the testing object is found.

## 2.3 Simplified Two-Dimensional PCA

### Training

The dimensions of input images are reduced by down-sampling according to the parameter $r$, in the same manner, as in the 1DPCA case. The simplified 2DPCA-based recognition is performed on a special (line-based) representation of images [12]. This is a difference form the full two-dimensional approach presented in [8]. For a grayscale images, the input to the training stage are $M$ vectors (lines) of size $1 \times N$, while for color images, on the input to the 2DPCA training stage are matrices of size $MN \times D$ elements (we perform a concatenation of triples consisting of pixel values in the respective color model. All image vectors are then added to the special four-dimensional structure, where the first 3 dimensions are responsible for the image dimensions and the last one is the image number. We use the simplified 2DPCA algorithm [12], which performs the following steps:

1. The whole database is employed at the stage of mean vector calculation, followed by normalization (through centering).
2. A covariance matrix of size $MN \times MN$ is calculated for training part of the dataset only, and decomposed into eigenvalues and eigenvectors.
3. According to the typical energetic approach, the $m$ most important eigenvectors are selected to form the transformation matrix of size $MN \times m$.
4. Finally, the training images are reduced to the $m \times D$ size, each.

**Testing**

The testing procedure involves adequate reduction and distance calculation (classification), similar to the one presented above.

## 3 Experiments

The dedicated software was created in Matlab. First, the input images (stored in JPEG files) are loaded and transferred into an appropriate color-space, then vectorized (in case of 1DPCA) or transformed into matrices (in case of 2DPCA). In order to make traditional PCA applicable and to limit the computation time, images are down-sampled according to the current test setup parameters. In order to find the most optimal color-space for the purpose of color image recognition we made several series of experiments involving different color representations, i.e. RGB, HSV, YIQ and YCbCr, as they are the most frequently used ones in the pattern recognition area.

The experiments involve the following recognition scheme. Training images are used to calculate transformation matrix. Then each training image is reduced - they are gathered in the reduced database used for classification. Above calculations are done in an off-line manner. The next part is on-line processing, which includes reduction of each test image and calculation of distances to all reduced training images. We use standard Euclidean metrics. The closest image in the training database points the class of the testing one.

In order to make the results more objective we employ cross-validation technique, since it is a very handy tool to protect against testing hypotheses suggested by the data. We performed 19 rounds of cross-validation, which involved random selection of training and testing images in each class. In our experimental scenario, the number of training images was increased from 1 to 19, while the number of testing images was decreased from 19 to 1, respectively. The tests was performed across different numbers of training/testing

**Table 1** Characteristics of databases used in the experiments

| Dataset | Faces94 | Faces95 | Faces96 |
|---|---|---|---|
| Number of classes ($K$) | 152 | 72 | 151 |
| Number of images in class ($Q$) | 20 | 20 | 20 |
| Total number of images ($KQ$) | 3040 | 1440 | 3020 |
| Controlled acquisition conditions | + | – | – – |
| Original image dimensions [px] | $180 \times 200$ | $180 \times 200$ | $196 \times 196$ |
| Cropped facial area size [px] | – | $120 \times 160$ | $120 \times 160$ |

objects and dimensionality of features after reduction in order to find the most optimal parameters for each algorithm.

## 3.1 Datasets

All algorithms were tested on publicly available datasets containing facial portraits: Faces94, Faces95 and Faces96. All datasets are provided by University of Essex, Department of Electronic Systems Engineering and have different number of classes, 20 images (24-bit RGB color JPEGs) each. All datasets present isolated faces with different expression, and changing shot orientation. Images contain portraits of male and female subjects, people of various racial origins, different age. Some of subjects wear glasses and have beards. Lighting is artificial, mixture of tungsten and fluorescent overhead. In case of Faces94 and Faces95 all faces are photographed over a uniform background, while in

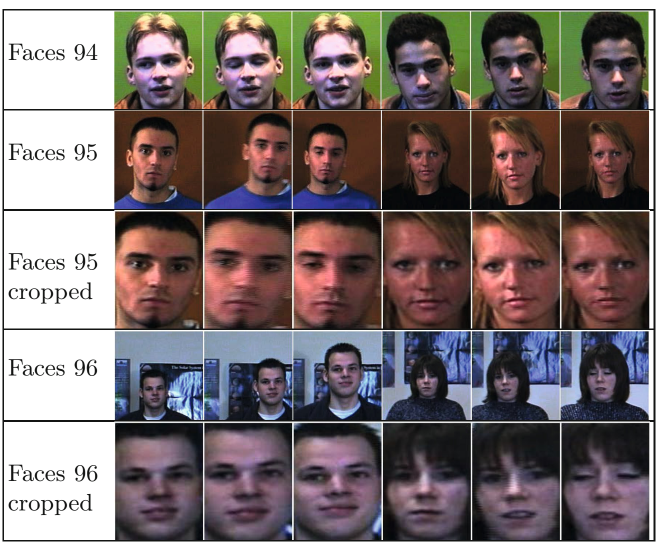



**Fig. 1** Several images from Faces Datasets used in the experiments

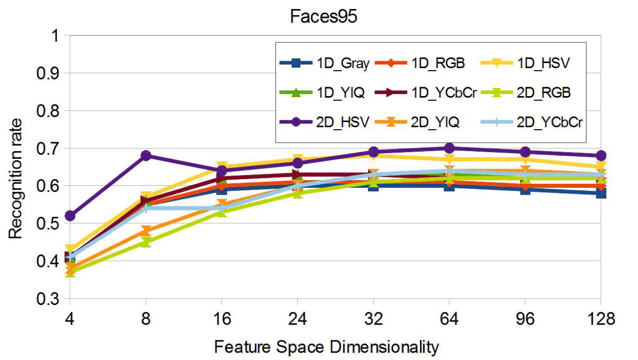



**Fig. 2** The results of recognition for Faces 95 database

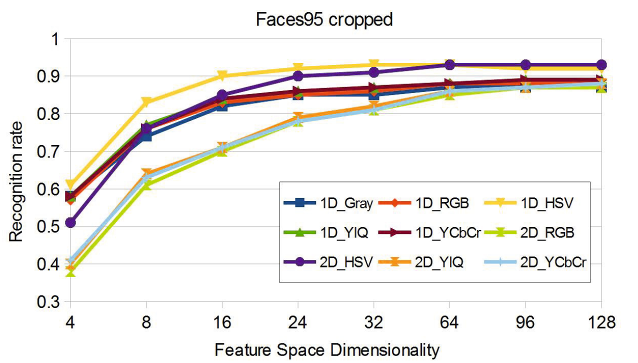



**Fig. 3** The results of recognition for cropped Faces 95 database

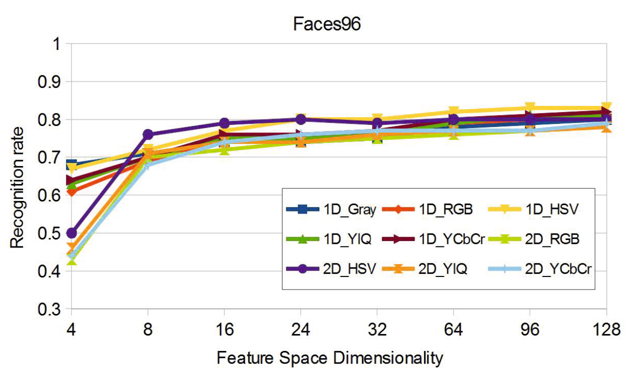



**Fig. 4** The results of recognition for Faces 96 database

case of Faces96, the images present people over some complex background. Faces occupy most of the picture. Detailed description of each database is given in Table 1. The experiments were performed on images taken directly from each set, as well as on cropped images, containing face-area only. The cropping was done in a semi-automatic way, using the distance between eyes (marked by human operator) as the reference dimension (see Fig. 1). The images were down-sampled to $26 \times 32$ in case of cropped faces and $32 \times 32$ in case of full frames.

### 3.2 Results

The results of experiments are presented in Fig.2 — Fig.5. The notation in the plots is as follows: first two symbols denote the transformation type (1D- or 2D-PCA), the other letters denote the color space. The accuracy is calculated as a ratio of the number of correctly classified images to the total number of test images. This ratio is an average value of results for experiments involving different number of testing and training images. As it can be seen, the simplified two-dimensional PCA is slightly superior to traditional one-dimensional PCA in the aspect of recognition accuracy in most cases. However, the two-dimensional transformation requires much less computations, since it employs smaller covariance matrix. It is an important advantage in case of color images. What is interesting, that the results for almost all color-representations are similar and generally higher than for grayscale representation. In many cases the highest recognition rate is obtained for HSV space, which is an important hint for practical implementations. The last interesting observation is that even simple cropping increases the recognition accuracy by more than 20 %. It should be also noted, that the experimental results for Faces94 are not shown, since this database is very standardized and the accuracy for all

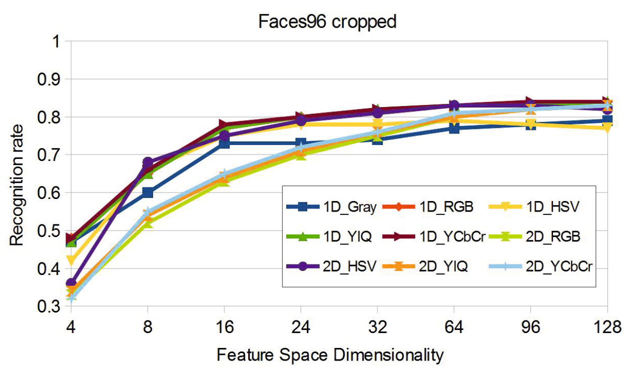



**Fig. 5** The results of recognition for cropped Faces 96 database

variants of color representation and dimensionality of PCA is close to 0.98, thus this database is not a challenging test-bed for such algorithms.

## 4 Summary

In the paper we presented a problem of reducing dimensionality of data structured in three-dimensional matrices, like true-color digital images. Unlike the cases reported in the literature the processing is done by means of simple two-dimensional PCA on images organized as two-dimensional matrices. The experiments performed on typical benchmark datasets containing human faces showed that the transformation of input data (3D to 2D and an adequate color transformation) gives the possibility to use it in the tasks of full-color image recognition with results higher than for traditional approach, without an extensive computational effort.

## References

1. Zuo, W., Wang, K., Zhang, H.: Subspace Methods for Face Recognition: Singularity, Regularization, and Robustness - State of the Art in Face Recognition. I-Tech, Vienna (2009)
2. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. ACM Computing Surveys 35(4), 399–458 (2003)
3. Jolliffe, I.T.: Principal Component Analysis. Springer, NY (1986)
4. Kirby, M., Sirovich, L.: Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans. PAMI 12(1), 103–108 (1990)
5. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
6. Tsapatsoulis, N., Alexopoulos, V., Kollias, S.: A Vector Based Approximation of KLT and Its Application to Face Recognition. In: Proc. of The IX European Signal Processing Conference, EUSIPCO 1998, Island of Rhodes, Greece (September 1998)
7. Swets, D.L., Weng, J.: Using Discriminant Eigenfeatures for Image Retrieval. IEEE Trans. PAMI 18 (1996)
8. Kukharev, G., Forczmański, P.: Data Dimensionality Reduction for Face Recognition. Machine Graphics & Vision 13(1/2), 99–122 (2004)
9. Wendy, S., Yambor, B.A., Draper, J., Beveridge, R.: Analyzing PCA-based Face Recognition Algorithms: Eigenvector Selection and Distance Measures. Computer Science Department Colorado State University Fort Collins (2000)
10. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear Analysis of Image Ensembles: TensorFaces. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part I. LNCS, vol. 2350, pp. 447–460. Springer, Heidelberg (2002)

11. Lu, H., Plataniotis, K.N., Venetsanopoulos, A.N.: MPCA: Multilinear Principal Component Analysis of Tensor Objects. IEEE Transactions on Neural Networks 19(1) (2008)
12. Yang, J., Zhang, D., Frangi, A.F., Yu Yang, J.: Two-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Trans. PAMI 26, 13–137 (2004)
13. Borawski, M., Forczmański, P.: Gram-Schmidt Orthonormalization-Based Color Model for Object Detection. Information Processing and Security Systems, Part I, 87–94 (2005)

# Recognition of Hand Drawn Flowcharts

Wioleta Szwoch and Michał Mucha

Gdansk University of Technology
wszwoch@eti.pg.gda.pl, zerojedynkowy@gmail.com

**Summary.** In this paper the problem of hand drawn flowcharts recognition is presented. There are described two attitudes to this problem: on-line and off-line. A concept of FCE, a system for recognizing and understanding of freehand drawn on-line flow charts on desktop computer and mobile devices is presented. The first experiments with the FCE system and the planes for future are also described[1].

## 1 Recognition of Hand Drawn Diagrams

In the days of common use of computers in almost all domains of our life hand drawn sketching on paper or electronic tablets still remains the most natural way of expressing ideas in a symbolic form. This way is more convenient and effective because a hand is more efficient and precise when using a pen then a mouse and keyboard.

That is why algorithms for recognition of hand drawn technical images are developed and automatic conversion of sketches into digital form has many applications. Some researches try to create a universal system that would deal with recognition of sketches of different kinds but most researches focus on specific applications such as UML diagrams, mechanics, architecture, handwritten text, handwritten music scores and so on [10, 4, 8, 11, 5, 2]. Researchers use different methods and attitudes to figures and text recognition e.g. Support Vector Machines (SVM) or Hidden Markov Models (HMM).

Flowcharts are very important for mathematicians, computer scientists. They allow to visually describing different algorithms in graphics form increasing their readability. Flowcharts are helpful during computer program creating and testing. They constitute work documentation.

In this paper a framework for recognition of hand drawn diagrams are proposed. This framework allows for on-line and off-line recognition. Its scalability allows for implementation on PC and mobile devices.

[1] This paper is partially sponsored by the Polish Government's research funds as a research project No DEC1-2011/01/B/ST6/06500.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 65–72.
springerlink.com 

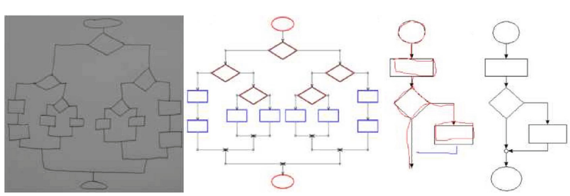
In section 2 general schema of diagram processing and characteristic of on-line and off-line attitudes are presented. Section 3 describes FCE - FlowChart Editor - system for on-line flowchart recognition. In section 4 first tests are presented. Section 5 presents conclusions and future works.

## 2 Method of Analysis Hand Written Documents

The general classification of methods of hand drawn documents analysis is involved with the way of graphical input into the computer system. In general there are two methods for graphical input to existing systems: on-line systems that allows for interactive input of information and off-line processed complete documents that was created earlier.

### 2.1 Off-Line and On-Line Attitudes

On-line and off-line attitudes to sketches analysis may differ in many aspects have a specific features, advantages and disadvantages. In on-line methods user creates drawings using electronic input device watching user movement such as a computer mouse or an electronic pen (tablet, tablet PC or a hand-held one). The information about painted objects is immediately available in the system as the list of subsequent pen or mouse coordinates. In such system no image preprocessing is needed and very simple segmentation algorithms may be used. Efficiency of such systems is very high because of high efficiency of segmentation and user interaction possibility. The number of devices with touch screen especially mobile grows rapidly then on-line attitude has a good perspective. In the off-line systems graphical images of paper drawings are processed. These images may be acquired by scanners or digital cameras. Unfortunately such images need special preprocessing and segmentation algorithms to be used. The aim of these algorithms is to extract and recognize the drawing objects from the input image. This task is not simple as input images



(a) (b) (c) (d)

**Fig. 1** Examples of off-line (FCA system) and on-line (FCE system) processing a) input image to FCA b) output image from FCA c) sketch and on-line recognized part of flowchart in FCE d) output flowchart from FCE may be of bad quality, may contain noise or other artifacts such as connected symbols. Examples of flowcharts are presented in fig. 1.

## 2.2 Processing of Technical Documents into Digital Form

In general, transformation of hand drawn documents into digital form consists of such stages as image preprocessing, segmentation, recognition, understanding and representation in digital format.

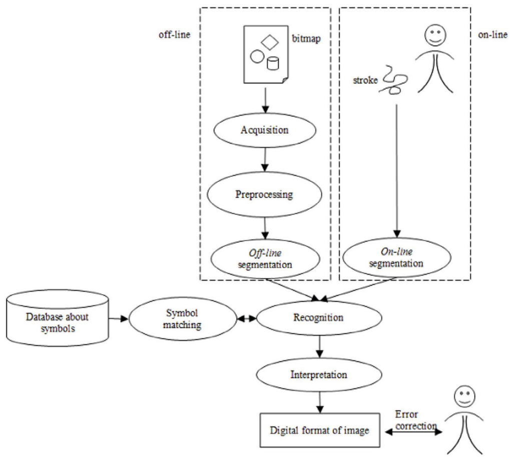



**Fig. 2** General schema on-line and off-line attitudes to technical documents processing

The general schema of transformation the technical documents into digital form is presented in fig. 2. The beginning stages of document processing depend on the kind of attitude used. In particular in off-line attitude the preprocessing is needed to improve image quality and prepare it for further operations which are segmentation and recognition. Preprocessing covers noise removing, binarization, thinning and vectorisation.

The goal of segmentation of technical drawings is symbols' separation from connections lines, basing on a set of general rules. In on-line attitude segmentation methods are simpler and more efficient.

The following processing stages are common for both attitudes (on-line and off-line). During recognition process detected segments are classified to one of allowed classes of symbols. In case of universal systems allowing for recognition of different kind of drawings information about shapes and symbol classes is placed in external knowledge database.

Understanding stage allows us know the structure of the image thanks its syntax and semantics analysis. Because the free hand drawings are usually careless, consisting figures of different sizes and without justification the aestheticization is very important issue. It allows for automatic optimization of their design, easier understanding and interpretation by people [7, 12, 1]. The final result is digital form of technical drawing that can be visualized and edited. Digital representation of the image allows for simple data exchanging between different applications and usage of drawings created in other applications. As automatic conversion not always is perfect there should be subsystem for error correction.

The most important problems during recognition of technical drawings are: imprecise of sketches, variety of symbols, overlapping of elements, bad quality of paper documents and incorrect of acquisition process.

## 3 System for On-Line Recognition of Hand Drawn Flowcharts

In order to confirm the correctness of proposed recognition framework the system FlowChart Editor (FCE) for on-line analysis was implemented. FCE allows to recognizing and understanding hand drawn flowcharts. Semantic analysis bases on proposed by the author graph grammar FlowGram [12]. Edition and error correction is allowed. FCE may also works as an usual flowchart editor building schema from predefined figures. FCE allows for parallel work as drawing and editing flowcharts. There is also possibility to store a recognized flowchart in proper format and use it in standard applications like Microsoft Word, PowerPoint etc.

### 3.1 Segmentation

Segmentation in on-line attitude is based on stroke recognition. There are two methods: primitive shape recognition (or single-stroke recognition) and composite shape recognition (or multi-stroke recognition) [9].

The main features that characterize on-line segmentation are: information about strokes (number, order etc.), possibility of user interaction and error correction just when they occur, accuracy and speed, less complicated segmentation algorithms and no preprocessing (as compared with off-line attitudes), ability discrimination of superimpose figures, need more complicated user interface.

Because flowchart's figures are relatively simple in FCE single-stroke recognition is used. It is not a heavy restriction for user, on the other hand such segmentation is easier and simpler, this method is fast and efficient, and not need additional segmentation and context analysis like during multi-stroke recognition [5]. Figures are recognized just after ending the drawing and may be redrawn in a more esthetic form.

Recognition module obtains such primitive (stroke) and is able to classify it to proper figures or connection class using simple rules.

### 3.2 Gestures, Figures and Connections Recognition

There are three recognition strategy in on-line attitude: continuous recognition - shapes are recognized continuously during drawing, algorithm try to look ahead object type, simultaneous recognition - strokes are recognized just after ending their drawing and recognize on demand (ROD) (or lazy recognition) - user decides when start recognition - it is useful when the symbol is complicated and the user determines symbol boundary [9, 5]. In FCE system simultaneous recognition and also recognition on demand were implemented.

#### Gestures

During on-line creation of drawings two kinds of interaction are available: a user makes gestures that means input of new diagram elements or gestures meaning commands for object manipulation. FCE recognizes several commands that are marking, deleting, moving and size changing.

Mobile version of FCE receives commands by messages generated by touch screen (touch, drop, untouched) and doesn't force the user to draw any special commands on screen. This solution is simpler and more intuitive and allows for a greater precision. Desktop version of FCE has similar solution using mouse. FCE examines possible interpretation of stroke and decides whether it is a symbol, connection or a command. Because of inaccurate drawing using computer mouse each stroke is smoothing.

#### Figures and Connections

The free hand drawings are usually careless and inaccurate - lines are not straight and not properly connected or even disconnected. The used algorithm for figures recognition is then very important for correctness of flowchart recognition. There are three the mostly applied types of algorithms: statistical - the statistical object features are evaluated, structural - symbol is described as set of geometrical primitives and spatial relationships among them, and also based on model [3, 10].

FCE uses pattern similarity method (SM) to classify figures. Each figure is compared with ideal and deformed patterns of each symbol class (Fig. 3). The number of deformed variants of patterns depends of individual geometric features and possible deformations in on-line drawings. The similarity is expressed by average value D of distances d determined for each point of a figure. A figure is classified to the class which base or variant pattern gives the smallest value D.

FCE offers also original DAS (Derivatives and Area Statistics) classification algorithm that bases on probabilistic analysis [13]. The classifier was

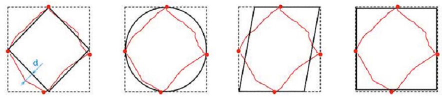


**Fig. 3** Example of distance between diamond and four basic figures

learned using a set of learning examples. Each figure is represented by a 6 geometric features such as number of piece linear element of a contour, contour's length to bounding rectangles length ratio and others. The number of features is enough to assure good separability of classes in 6-dimensional features space.

Connections are drawn as a single or multi strokes. Direction of control flow is determined automatically using context analysis [12].

### 3.3 Structure Understanding

In FCE system syntax of flowchart is veryfied based on a graph grammar FlowGram [12]. Graph grammars provide a mechanism in which generation and transformation of visual objects can be modeled precisely in a mathematical way. FlowGram graph grammar defines graphical graph rewriting rules, predicates of their applications and attributes evaluation methods.

The set of terminal symbols of FlowGram consists of 5 symbols: start/stop, instruction, i/o, decision and node. There are only two nonterminal symbols in FlowGram: block and the start symbol. Though FlowGram is context sensitive, there is no problem to construct a proper bottom-up parser according to its rules. The appliance order of graph grammar rules is very important. For FlowGram that order is specified by the grammar programming graph.

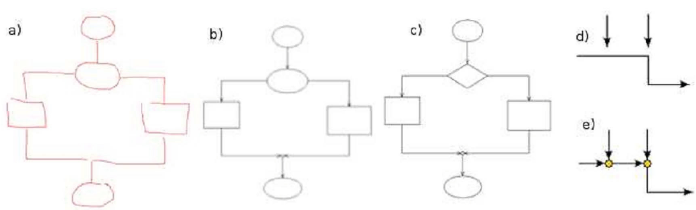



**Fig. 4** Examples of contextual analysis in FCE a) drawn flowchart b) recognized flowchart c) flowchart after contextual analysis d) e) automatic join of connections

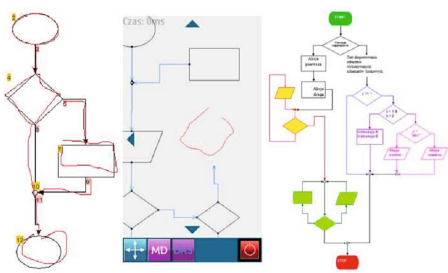
FCE system does context analysis of each stroke just after its ending (Fig. 4). The system corrects symbol location, joins symbols and connections, automatic creates nodes, recognizes element type. In FCE contextual analysis of complete flowchart is enabled using validator.

## 4 Experimental Results

To validate the recognition quality of the FCE system experiments were carried out on 120 handwritten figures. During experiments computer with Pentium Dual-Core 2,6GHz and Windows 7 was used. For the testing images the recognition quality varied from 90,8% for SM algorithm up to 97,5% for formulated DAS algorithm (tab. 1). It is important that recognition of flowcharts gives better results and have better efficiency than figures recognition because of context analysis. Such analysis allows for automatic correction some kind of errors like for instance recognition ellipse as other figures. The average processing time of DAS algorithm for single figure was about 0,06ms. FCE mobile application was tested on Nokia C5-03 phone.

**Table 1** Efficiency of figure recognition for DAS algorithm

| figure type | number | rectangle | parallelogram | rhombus | ellipse | errors | efficiency |
|---|---|---|---|---|---|---|---|
| rectangle | 30 | 28 | 2 | | | 2 | 93,3% |
| parallelogram | 30 | | 30 | | | | 100% |
| rhombus | 30 | | | 30 | | | 100% |
| ellipse | 30 | | 1 | | 29 | 1 | 96,67% |
| all | 120 | | | | | 3 | 97,5% |




(a) (b) (c)

**Fig. 5** Example of FCE system a) sketch and recognized flowchart b) FCE on mobile devices c) flowchart exporting from FCE to Microsoft World

## 5 Conclusions

In this paper the problem of analysis of documents containing flowcharts was presented. The possible attitudes to this problem were described. Presented prototype FCE system uses on-line attitude and may be used on desktop computers and mobile devices. Tests of its efficiency were also described.

FCE system allows for easy and intuitive hand drawn creation of flowcharts and converting them into a standardized digital form. It is compliant with nowaday trends of developing applications based on intelligent systems that actively interact with a user. Proposed framework and applications may be used in many applications. Its practical use is assured by possibility of diagrams exporting in popular office formats such as .docx, .pptx, .odt, .odp.

Now two separate systems FCE and FCA enable on-line and off-line analysis of flowchart. Further works will focus on joining these systems and developing a module for automatic generation of source code in programming language. Prototype of such modul was tested in FCA - off-line version of the system. The proposed framework was verified for a large class of flowcharts. Further works will focus on applying of this framework for other types of technical drawings.

## References


1. Szwoch, M.: Formal languages, automata and translators. PWNT, Gdańsk (2008) (in Polish)
2. Cao, H., Prasad, R., Natarajan, P.: Handwritten and Typewritten Text Identification and Recognition using Hidden Markov Models. In: Int. Conf. on Document Analysis and Recognition ICDAR, pp. 744–748. CPS IEEE Computer Society (2011)
3. Lin, Z., He, J., Zhong, Z., Wang, R., Shum, H.Y.: Table Detection in On-line Ink Notes. IEEE Transactions on PAMI 28(8), 1341–1346 (2006)
4. Blostein, D., Lank, E., Thorley, J., Chen, S.: On-line Recognition of UML Diagrams. In: Sixth Int. Conf. on Document Analysis and Recognition ICDAR, pp. 356–360. CPS IEEE Computer Society (2001)
5. Kara, L.B., Stahovich, T.F.: Hierarchical Parsing and Recognition of Hand-Sketched Diagrams. In: 17th ACM Symp. on User Interface Software and Technology (2004)
6. Szwoch, W.: Recognition, understanding and aestheticization of freehand drawing flowcharts. In: ICDAR 9th Int. Conf. on Document Analysis and Recognition, Brazil, vol. 2, pp. 1138–1142. CPS IEEE Computer Society (2007)
7. Szwoch, W.: Aestheticization of Flowcharts. In: Stapleton, G., Howse, J., Lee, J. (eds.) Diagrams 2008. LNCS (LNAI), vol. 5223, pp. 423–426. Springer, Heidelberg (2008)
8. Wang, X., Xu, H., Wang, H.: On-line Sketch Recognition for Course of Action Diagrams. In: IEEE Int. Conf. on Mechatronics and Automation, pp. 465–469 (2010)
9. Liu, W.: On-line Graphics Recognition: State-of-the-Art. In: Lladós, J., Kwon, Y.-B. (eds.) GREC 2003. LNCS, vol. 3088, pp. 291–304. Springer, Heidelberg (2004)
10. Yu, Y., Samal, A., Seth, S.C.: A system for recognizing a large class of engineering drawings. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(8), 868–890 (1997)
11. Hammond, T., Paulson, B.: Recognizing Sketched Multistroke Primitives. ACM Transactions on Interactive Intelligent Systems, 1–34 (2011)
12. Szwoch, W.: Automatyzacja procesu analizy odręcznych schematów blokowych i oceny ich estetyki. PhD Thesis, Gdansk University of Technology, Gdańsk (2009)
13. Mucha, M.: Rozpoznawanie ręcznie rysowanych schematów blokowych. MA Thesis, Gdansk University of Technology, Gdańsk (2011)

# Mean Shift Based Automatic Detection of Exudates in Retinal Images

Juan Martin Cárdenas[1], M. Elena Martinez-Perez[2], Francesc March[3], and Nidiyare Hevia-Montiel[2]

[1] Engineering Postgraduate Faculty, National Autonomous University of Mexico (UNAM)
juan.cardenas@comunidad.unam.mx
[2] Department of Computer Science, Institute of Research in Applied Mathematics and Systems (IIMAS), National Autonomous University of México (UNAM)
elena@leibniz.iimas.unam.mx, nidiyare@turing.iimas.unam.mx
[3] Instituto de Oftalmologia. Fundacion de Asistencia Privada Conde de Valenciana IAP. Mexico, DF
marchfrancesc@gmail.com

**Summary.** Exudates are one of the principal lesion present in the normal development of Diabetic Retinopathy (DR), its detection is an important step in (DR) screening and classification. This paper presents an automated method for bright lesions detection in retinal images by means of the mean shift filtering. Due to uneven illumination of retinal images it is necessary to perform a preprocessing step consisting of a shade correction technique finding non-structures pixels and adjusting a third order polynomial to be substracted from the original image. The mean shift filtering is applied to enhance bright areas and to uniform background non-structures regions. A region growing algorithm is performed from local maxima regions taken as seeds to get the final results. A set of 20 retinal images selected and manually tagged by a retinal specialist ophthalmologist were used for the evaluation. Results present a true positive rate (TPR) of 0.627 and a specificity SPC of 0.979. It is demonstrated that Mean shift filtering is a promising method for exudates detection.

## 1 Introduction

Diabetes Mellitus (DM) has became a public health problem worldwide. Diabetic Retinopathy (DR) is the damage that can be regarded as manifestation of diabetes on the retina. DR can eventually lead to blindness; this risk could be reduced if there is an early detection and proper treatment and monitoring of the eyes. Screening and classification to detect DR include different retinal manifestations; one of the most important is the presence of hard exudates (HE). HE are extracellular accumulations of lipoproteins derived from leakage from abnormal vessels; are a marker for the presence of retinal oedema. HE have a white-yellowish appearance and often with circular pattern around broken vessels [11]. Distribution and location of HE are relevant in clinical

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 73–82.
springerlink.com 

approach, specially if hard exudates are in the macula, the main area responsible of vision. Detecting HE lesions is very expensive in professional time and automated screening could offer a preliminar evaluation of the patients in diagnosis and follow-up, and valuable information for studies.

## 1.1 Exudates Detection

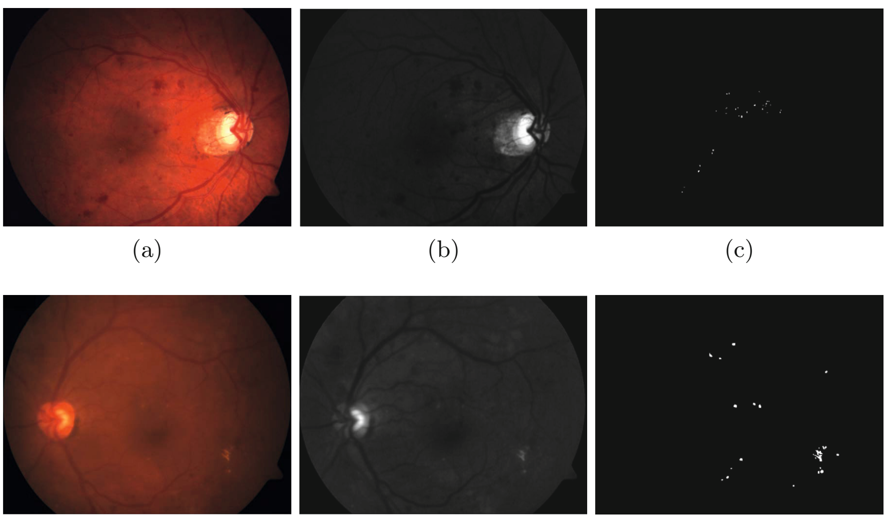
Several papers have presented solutions to computer based exudates detection. In [9] the green channel is used to apply a morphological closure with an octagonal structuring element to enhance the exudates-background contrast. They propose a ratio between the addition and substraction of the green and red channel and the exudates are found by a thresholding. [12] propose a statistical solution based on mixture models (MM) to separate HE from the background. In [10] a fuzzy c-means clustering algorithm is proposed. First, a pre-processing step consisting of color normalization and contrast enhancement is done. Next a whole segmentation between exudate-non exudate segmentation is performed using a set of features like color, size, edge strength and texture by means of a Multilayer Neural Network. In [2] the marker controlled watershed transformation is proposed to exudates detection. First, a smooth filter is applied and the green channel is extracted from RGB space to apply a gamma transform. Two types of markers are used: external (associated to the background) and internal (associated to the object of interest). Internal markers are located from the image gradient using the extended minima transformation and external markers are located by partitioning the image in order to maintain one internal marker inside each external region. [6] propose a solution to white lesions detection based on pure splitting adaptive threshold algorithm. This solution is divided in four steps: 1) a pre-processing step by illumination correction in the green channel, 2) optic disc elimination, 3) pure splitting thresholding algorithm and 4) coarse segmentation. [8] propose an exudate enhancement by diffuse morphology. The image is converted to gray scale followed by a series of duffusse opening and closing operations. The resulting image is added to the original image obtaining a retinal image with bright lesions enhanced.

In this paper we propose a method for automatic exudates detection applying a mean shift filtering [1]. First, a pre-processing step for non-uniform illumination correction on the green channel is performed using a polynomial fitting from a set of background pixels (pixels different that vessels, optic disc, macula, and lesions such as microaneurysms, exudates, etc), next the mean shift filtering and segmentation is applied with empirical parameter selection to reduce the number of gray levels. A region growing algorithm is applied using seeds obtained by a local maxima selection to generate the final binary result. Results are evaluated using a *ground truth* images and sensitivity and specificity is calculated.

## 2 Materials and Methods

A set of 20 retinal images with considerable non-uniform illumination and with presence of white lesions was selected. The images were captured using a Carl Zeiss FF450 plus fundus camera, acquired in JPEG format with 2588 × 1958 pixels resolution. The original images and the *ground truth* images were resized to 647 × 489 pixels to reduce the overall time processing using bi-cubic interpolation and nearest neighbor interpolation respectively. The entire process is performed on the green channel.

The *ground truth* images were manually outlined by a retinal specialist (FM) using a tool designed for that purpose. The tool was written in Matlab with basic capabilities of zoom, polygon drawing, erase and cut. The specialist was advised to delineate as close as possible every white lesion (hard and soft exudates) present in each image. The mean time of delineation was 15 min per image. Figures 1(c) and 1(f) show two examples of the *ground truth* images.



(a) (b) (c)

(d) (e) (f)

**Fig. 1** (a) and (d) Original images, (b) and (e) shade corrected images respectively and (c) and (f) manually outlined template by an specialist respectively

### 2.1 Non-uniform Illumination Correction

We follow the same additive image model proposed in [3] in which the observed funds image $I$ can be modeled as:

$$I = f(I^0) = f(I_b^0 + I_f^0) \tag{1}$$

where $I^0$ is the original image, $I_b^0$ is the original background image and $I_f^0$ is the original foreground image and $f(\cdot)$ is the acquisition transformation. $I_f^0$ is the image with all structures (vessels, optic disc and macula) and lesions (microaneurysm, exudates, etc) and $I_b^0$ is an image without any structure or lesion. In this manner, the corrected image or the original foreground image can be calculated by means of:

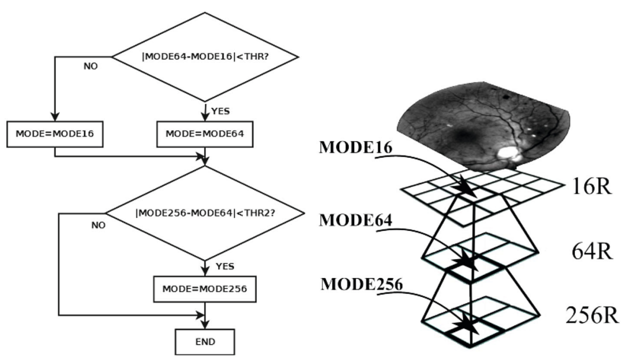
$$I^0 - I_b^0 = I_f^0 \tag{2}$$

Our non-uniform illumination correction procedure is divided in three steps:

1. Extraction of background pixels.
2. Adjustment of a third order polynomial.
3. Shade corrected image result by substracting.

The main idea of background pixels extraction is based on the observation that principal structures and lesions presented in the image are smaller than non-structures and non-lesions areas so the mode of a gray level histogram corresponds usually to the background pixels over a small window.

The image is tessellated in 16 , 64 and 256 regions (4 by 4, 8 by 8 and 16 by 16 rectangular regions called 16R, 64R and 256R respectively). Each 16R region has 4 64R regions associated as well as each 64R region has 4 256R region. A mode (most frequent gray level) is calculated for each 16R, 64R and 256R region. If the mode of the 64R region is inside ±10% of the 16R region mode the 64R mode is selected, else the 16R associated region mode is selected. In the same way, if the mode of the 256R region is inside ±10% of the 64R region mode the 256R mode is selected, else the 64R associated region mode is selected. Figure 2 shows the process of background pixel set selection.




(a) (b)

**Fig. 2** Process of background pixel set selection. (a) schematic diagram of comparison sequence and (b) representation of tessellation and mode selection step.

We choose one background pixel per 256R region as that who have the same gray level that the mode selected in each 256R region and that one which does not belongs to an edge by comparing the gradient magnitude with a fixed threshold. This give us an almost equally spaced grid of pixels that belong to the background image. This structure yields an equally weighted surface fitting in the next step, otherwise there could be areas with more influence than other ones (e.g. borders).

A 3 order polynomial surface of the type:

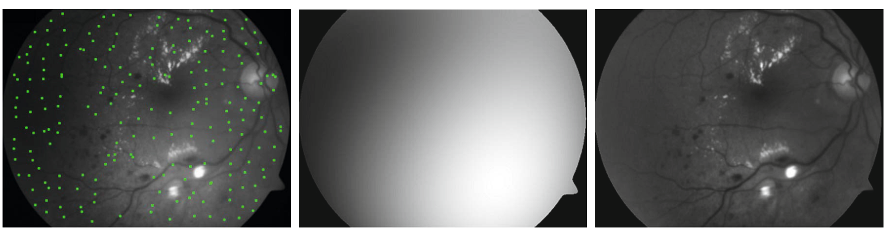
$$a_0 + a_1x + a_2y + a_3x^2 + a_4y^2$$

$$+a_5xy + a_6x^3 + a_7x^2y + a_8xy^2 + a_9y^3 \tag{3}$$

is adjusted using least squares over the 256R grid of selected background pixels. This surface is adopted as an estimation of the original background image $I_b^0$. Finally, the original foreground image is calculated by substracting in the form

$$I_f^0 = I - I_b^0 \tag{4}$$

Figures 1(b) and 1(e) shows two results of the non-uniform illumination correction algorithm. Figure 3 shows the entire process.



(a) (b) (c)

**Fig. 3** Non-uniform illumination correction Process. (a) Original green channel image with the point selection grid, (b) Third order polynomial surface fitting by least squares and (c) final shade corrected image.

## 2.2 Mean Shift Filtering and Segmentation

The mean shift (MS) algorithm [4] is a nonparametric method of kernel estimation to cluster a set of observations into different classes, by assigning each observation to the nearest mode along the direction of the gradient at the observation points. Let $f(x)$ be the unknown probability density function (PDF) underlying a d-dimensional Euclidean space, and $X_i$ the available data points in this space. The MS simplest formulation can be written as in [7]:

$$\hat{\nabla} f(x) \sim M(x) = \frac{1}{n_x} \sum_{\mathbf{X}_i \in S_h(x)} \mathbf{X}_i - x \qquad (5)$$

where $S_h(x)$ is the $d$-dimensional hypersphere with radius $h$ centered on $x$ and containing $n_x$ data points. The MS vector (5) can define a path leading $x$ to a stationary point (that is, to the *modes*) of the estimated PDF.

As stated in [5], the most frequently used edge detection methods are based on gradient orientation. An improved MS estimate can be obtained by weighting each pixel within the region by a function of its edge confidence ($\varphi$), so that voxels that lie close to an edge (edge confidence $\approx 1$) are less influential in the determination of the new cluster center. The modified MS that includes weighted edge confidence is, from eq.(5):

$$M(x) = \frac{1}{\sum (1 - \varphi_i)} \sum_{\mathbf{X}_i \in S_h(x)} (1 - \varphi_i)\mathbf{X}_i - x \qquad (6)$$

An Euclidean metric is used to control the quality of the segmentation, which is dependent on the radii $h_s$ and $h_i$, corresponding to the resolution parameters of the estimate in the spatial and intensity (range) domains. After the MS procedure is applied to each data, those points that are sufficiently close in the joint domain are fused to obtain the homogeneous regions in the image. The number of clusters present in the image is automatically determined by the number of significant modes detected.

In order to fuse together homogeneous adjacent regions that have been split apart by the MS procedure, and to integrate the discontinuity information into the fusion step, the transitive closure operations are then performed on the region adjacency graph (RAG). Figure 4(a) shows the MS segmented image applied to figure 1(b).

### 2.3 Maxima Segmentation

After the MS segmentation is done there is still a large number of regions in the image thus a standard local maxima finding is performed followed by a region growing algorithm. First, a maxima and minima filter is applied to the MS segmented image with a window size of $7 \times 7$ pixels (heuristic selection). The binary mask images $I_{max}$ and $I_{min}$ are obtained as those pixels that have the same intensity in the MS segmented image as the maxima and minima filtered image respectively. At this moment, the optic disc (OD) is eliminated by hand, further work is been doing in order to eliminate it automatically. The maxima seeds pixels are chosen as those pixels that belong to the $I_{max}$ image and that do not belong to $I_{min}$ and to the OD area. Those pixels are taken as seed for a region growing algorithm that will finally result in the HE regions.

In the region growing algorithm each seed is taken as a different class. The process is done by iteratively dilating the seed region until its size has not changed. At every iteration each pixel of the region is compared with the mean of the class and is accepted if the distance is not grater than a fixed threshold (empirical value set up to 10), in other words, after each dilation the pixels added to the region are accepted if:

$$|x_i - \overline{c_i}| \leq 10 \tag{7}$$

where $x_i$ is the added pixel intensity in the MS segmented image and $\overline{c_i}$ is the actual mean of the class. The resulted image is 3 times dilated with a $3 \times 3$ pixel rectangular structuring element.

Figure 4(b) shows the seeds found from figure 4(a) and figure 4(c) shows the result segmentation.

(a) (b) (c)

**Fig. 4** HE segmentation process. (a) Image segmented with MS algorithm, (b) seeds found and (c) final result after maxima finding segmentation and region growing.

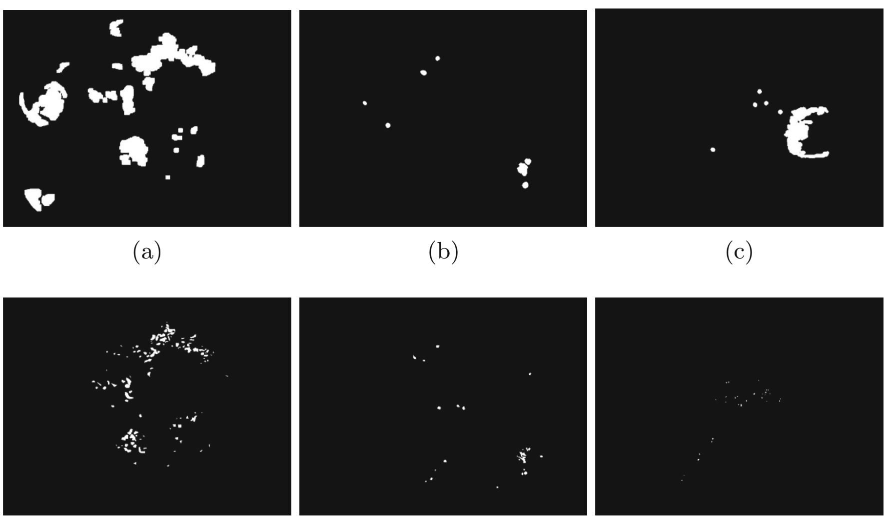
## 3 Results

A numerical evaluation was performed by comparing the results obtained by the proposed algorithm and the *ground truth* images. Sensitivity or true positive rate (TPR) and Specificity (SPC) were calculated following the ROC analysis for each image, the results are shown in Table 1, mean values for 20 images where TPR=0.627 and SPC=0.979.

First row of figure 5 shows three different performance results with the corresponding *ground truth* images shown in the second row. Surrounding areas of the OD and the surrounding areas of the main vessels often become distractors for the segmentation algorithm and affect the final results. In some cases, the the algorithm is not capable to find several HE because the color and contrast difference is not so strong and they are mistaken with the background.

**Table 1** Sensitivity (TPR) and Specificity (SPC) of the automated detection

| IMAGE | TPR | SPC | IMAGE | TPR | SPC |
|---|---|---|---|---|---|
| Im01 | 0.873 | 0.980 | Im02 | 0.326 | 0.990 |
| Im03 | 0.720 | 0.982 | Im04 | 0.309 | 0.970 |
| Im05 | 0.602 | 0.993 | Im06 | 0.938 | 0.930 |
| Im07 | 0.731 | 0.981 | Im08 | 0.760 | 0.997 |
| Im09 | 0.722 | 0.988 | Im10 | 0.394 | 0.989 |
| Im11 | 0.767 | 0.995 | Im12 | 0.555 | 0.991 |
| Im13 | 0.482 | 0.992 | Im14 | 0.510 | 0.984 |
| Im15 | 0.757 | 0.959 | Im16 | 0.719 | 0.961 |
| Im17 | 0.791 | 0.972 | Im18 | 0.547 | 0.975 |
| Im19 | 0.423 | 0.971 | Im20 | 0.615 | 0.978 |
| **MEAN** | 0.627 | 0.979 | | | |



(a) (b) (c)

(d) (e) (f)

**Fig. 5** Segmentation result of 3 different performance images. (a), (b) and (c) are the detection results from images Im06, Im12 and Im04 respectively. (d), (e) and (f) are their corresponding *ground truth* images.

## 4 Conclusion

The MS segmentation is a powerful tool for detecting HE in retinal images. The proposed algorithm reaches a high SPC and an acceptable TPR.

The OD segmentation is an important step in the HE detection because of the similarity in color and intensity and the region growing algorithm has empirical parameters that minimize the robustness and flexibility. Work is being done in automatic parameter selection and OD automatic segmentation.

Comparisons with other methodologies, as those mentioned in Section 1.1, are difficult to perform since each group uses its own database. The idea is to have a public database available to all, and so compare the performance of each method published. Work is being done in this issue.

**Acknowledgement.** Special thanks to the Ophtalmology Institute *Fundación de Asistencia Privada Conde de Valenciana IAP* and the Corneal Department head, Dr. Enrique Graue.

This work was supported by "‘Consejo Nacional de Ciencia y Tecnología - CONACYT"’ by a national research grant/scholarship with CVU 376445.

## References

1. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), 603–619 (2002)
2. Eswaran, C., Reza, A.W., Hati, S.: Extraction of the contours of optic disc and exudates based on marker-controlled watershed segmentation. In: International Conference on Computer Science and Information Technology (2008)
3. Foracchia, M., Grisan, E., Ruggeri, A.: Luminosity and contrast normalization in retinal images. Medical Image Analysis 3(9), 179–190 (2005)
4. Fukunaga, K., Hostetler, L.D.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory IT-21, 32–40 (1975)
5. Hevia-Montiel, N., Jiménez-Alaniz, J.R., Medina-Banuelos, V., Yanez-Suárez, O., Rosso, C., Samson, Y., Baillet, S.: Robust nonparametric segmentation of infarct lesion from diffusion-weighted MR images. In: Proceedings of the 29th Annual International Conference of the IEEE EMBS (2007)
6. Jaafar, H.F., Nandi, A.K., Al-Nuaimy, W.: Detection of exudates in retinal images using a pure splitting technique. In: 32nd Annual International Conference of the IEEE EMBS, Buenos Aires, Argentina, (2010)
7. Jiménez-Alaniz, J.R., Medina-Banuelos, V., Yanez-Suárez, O.: Datadriven brain MRI segmentation supported on edge confidence and a priori tissue information. IEEE Transactions on Medical Imaging 25(1), 74–83 (2006)
8. Mansoof, A.B., Khan, Z., Khan, A., Khan, S.A.: Enhancement of exudates for the diagnosis of diabetic retinopathy using fuzzy morphology. In: IEEE International Multitopic Conference, INMIC (2008)
9. Mir, H., Al-Nashash, H.: Assessment of retinopathy severity using digital fundus images. In: 2011 1st Middle East Conference on Biomedical Engineering, MECBME (2011)

10. Osareh, A., Shadgar, B., Markham, R.: A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. IEEE Transactions on Information Technology in Biomedicine 13(4), 535–545 (2009)
11. Pereira Delgado, E.: Nuevas perspectivas en oftalmología: Retinopatía diabética. Glosa, Laboratorios Esteve (2005)
12. Sánchez, C.I., Mayo, A., García, M., López, M.I., Hornero, R.: Automatic image processing algorithm to detect hard exudates based on mixture models. In: Proceedings of the 28th IEEE EMBS Annual International Conference, New York City, USA (2006)

# Hybrid Algorithm to Human-Face Detection, Recognition and Unrestricted Tracking

Hector Vargas, Esperanza Medina, Daniel Martinez,
Edson Olmedo, and Gerson Beristain

Universidad Popular Autonoma del Estado de Puebla, Computer Faculty, UPAEP,
21 Sur 1103CP 72410, Puebla, Mexico;
{hectorsimon.vargas,aldodaniel.martinez,edson.olmedo,
gerson.beristain}@upaep.mx esperanza.medina@upaep.edu.mx

**Summary.** This paper uses different algorithms to build a hybrid system for detecting human-face and tracking unrestricted. The system uses the face detection algorithm, Kalman filter [1]. The architecture is as follows: it locates the face in the image and get a sub image from the region of the head, face patterns are determined as the eyes, the center of the face, the border of the head, these parameters are used in the Kalman filter to takes the final decision on the direction in which the face in the image and reduce the error when more than one person in the picture, especially when there is no face but we know that still another position. In face recognition [2], the algorithm takes the detection phase, cutting image of detected face, which is divided in 9 subsections [6], where histogram comparison process [8] and phase correlation are made [7], where given results are processed by a decision tree which makes the decision if face is known or not. The experimental results show that the system is stable when it is saturated field of view with many faces or people.

## 1 Introduction

One of the great challenges in computer vision is to improve the automated systems for objects detection and tracking or regions of a set of images. Nowadays, detecting human-face is a discussed problem using different types of features. Face tracking is a difficult problem because faces are deformable objects with areas of little texture. Many algorithms implement the face-detection task as a binary pattern-classification task. In other words, the content of a given part of an image is transformed into features, after which a classifier trained on example faces decides whether that particular region of the image is a face, or not. The process of face detection consists of two steps. The first is building models. The second step is to find a particular region in the image, called area of interest.

Another category of tracking algorithms takes advantage of correlations between image frames to accelerate tracking process. Basically, the features

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 83–93.
springerlink.com 

of the object itself are local information, and the features of an image sequence belong to global information.

Many works have been conducted for object detection and tracking. Most available algorithms focus on estimating movements of area of interest using probabilistic theories. Some popular models and approaches, like Kalman filter. Some researchers proposed different control and noise models [3] into the recursion function; however those assumptions are dependent on specific applications and need to be tuned carefully.

One example of a face tracking algorithm can be the one proposed by Baek [1]. The state vector of a face includes the center position, size of the rectangle containing the face, the average color of the face area and their first derivatives. The new candidate faces are evaluated by a Kalman estimator. In tracking mode, if the face is not new, the face from the previous frame is used as a template. The position of the face is evaluated by the algorithm.

In parallel, another point of attention has been the faces recognition. In this context is presented problems in the variability of the head rotation, the intensity and angle of light, facial expression to take a picture and others which have complicated the recognition systems [2, 9].

Face recognition methods and algorithms commonly assume [10] that face images are aligned and have an equal posture, but in many practical applications it is impossible that these conditions will be present. One solution that has been presented to reduce these factors is to obtain sub-features of a face, which enables a better approximation to say whether a face is known or not based on the analysis results and related features.

Some work on face recognition [9, 10] have shown that an adequate solution to measure the characteristics is the use of decision trees, a probabilistic technique that through certain events and probability weights can make a decision in context.

This paper is organized as follow. In section 2, related algorithms that will be used to built a hybrid system for detecting human-face detection, recognition and tracking. Section 3 describes the detailed information of hybrid system that we propose. The results are presented in section 4 and the conclusions in section 5.

## 2 Related Algorithms to Build a Hybrid System

To build a hybrid system for detecting human-face and tracking unrestricted is necessary to use different algorithms. In This paper we propose an algorithm for recognizing and tracking a person.

The goal to this algorithm is to search a face in an image using a recognition process based on the detection of features that encode some information about the face to be detected, once face image detected is cropped, the segmentation is to determine if a face is already known, continues to track the face, so it will trace the journey that the faces will be making and storing a image is taken from the coordinates in which the face is detected, it is able to continue to monitor if the face is not fully facing the camera.

When the face is back facing the camera turns to make the recognition to ensure that the face is the same as it began to follow, Fig. 1.

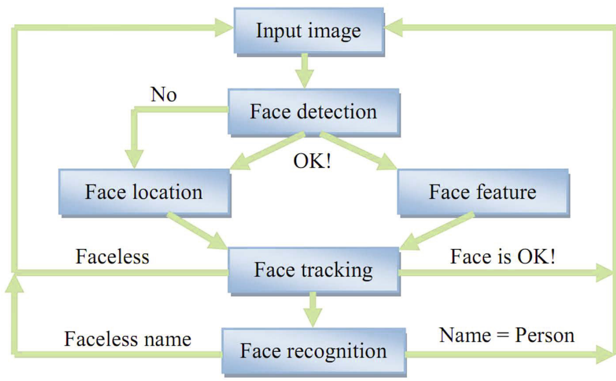



**Fig. 1** Hybrid system to Detection, Recognition and Tracking Unrestricted

Importantly, the algorithm must first find a face to follow the hybrid algorithm. We should also mention that this algorithm was tested on the service robot "Donaxi@HOME".

## 2.1 Face Detection

Face detection is the first stage of an automated face recognition system, since a face has to be located before it is recognized. In this case, we use the object detector of OpenCV. This object detector has been initially proposed by Paul Viola and improved by Rainer Lienhart. First, a classifier are trained with a few hundreds of sample views of a particular object (in this case a face), called positive examples, that is scaled to the same size (say, 20×20), and negative examples - arbitrary images of the same size.

After a classifier is trained, it can be applied to a region of interest (of the same size as used during the training) in an input image. The classifier outputs a "1" if the region is likely to show the object (face) and "0" otherwise. To search for the object in the whole image one can move the search window across the image and check every location using the classifier. The classifier is designed so that it can be easily "resized" in order to be able to find the objects of interest at different sizes, which is more efficient than resizing the image itself. So, to find an object of an unknown size in the image the scan procedure should be done several times at different scales.

### 2.2 Face Location

Once a face is detected within the image, the center coordinates of the face to know are stored so the system can map the exact location of the face within the image when the face changes its position; the new coordinates are also stores in order to know the person trajectory.

With the obtained data is generated a line graph as this type of graphic is used to compare values over time.

After we are going to analyze the behavior of the graph and discard those values that are far away from others to ensure that the plotted points show the path that the person made.

To find the direction and extent to which the person moved draw a line from the point where the face was detected for the first time to the last point where the face was detected. The information that we obtain after to detect a face is used in the Kalman filter configuration to face tracking.

### 2.3 Kalman Filter

In order to keep the person even when the face isn't fully in front of the camera we use the Kalman filter. Kalman filter estimates the state variables of a process with feedback. Calculate the process state at some point and then get information (fed back) of the measure.

If you want to apply the Kalman filter to face tracking, you must provide a characteristic representative thereof, to be taken as the observation of the object. In this case an image is stored each time a face is detected or a picture of where the face was detected last. To calculate this point we will call the center of mass and determines the position of the stored image is necessary to perform a series of operations on the image.

To identify movement of the face within the image is necessary to apply motion detection techniques. This way we continue to face that hovers over the image.

The study of movement of the face within the image is essential to introduce the time variable. A sequence of images is given by the function $f(x, y, t)$, where $x$, $y$ are the coordinates of the center of the head at a particular time instant, t. Therefore the value of $f(x, y, t)$ represents the intensity of pixel $(x, y)$ within the image t.

In face tracking measures are necessary observation corresponds with the position of center of mass of the face. In this case it is used to determine the position based on projections which use the center of mass, which, as mentioned above is obtained from the center point of the head. Previously eliminates potential noise given only to those parts of the image over a fixed number of pixels above the threshold. In this way we make sure not to consider some isolated pixel face.

### 2.4 Probabilistic Decision Trees

Decision trees [4] are a widely known formalism for expressing classification knowledge. The traditional approach to constructing a decision tree from a training set of cases described in terms of a collection of attributes is based on successive refinement. In a general context, the decision trees are particularly useful when one or more decisions in sequence should be taken and they are affected by one or more uncertain events to which they are assigned a certain probability weighting which directly affect the value (usually expressed in earnings) for final decision.

In the context of this work, we take the principles of decision trees for use in face recognition, using as a key decision if the face is known or not, and as uncertain events similarities and differences in the detected face real time, against the face stored in the database of known persons. Similarity values and differences were obtained using phase correlation algorithms and comparing histograms.

In order to achieve this aim with better efficiency, face division by subsections was used, based on the aesthetic theories [5] where each subsection is analyzed in an independent way with the similarity and difference algorithms.

Each one of these subsections received a probabilistic weight based on how characteristic for the face recognition is, for finally integrate them in a probabilistic decision tree.

### 2.5 Phase Correlation

It is a method for determinate the displacement of one picture respect to other similar [7]. However phase correlation utilizes a direct way comparison, results are too ambiguous because itŠs required that the compared face images

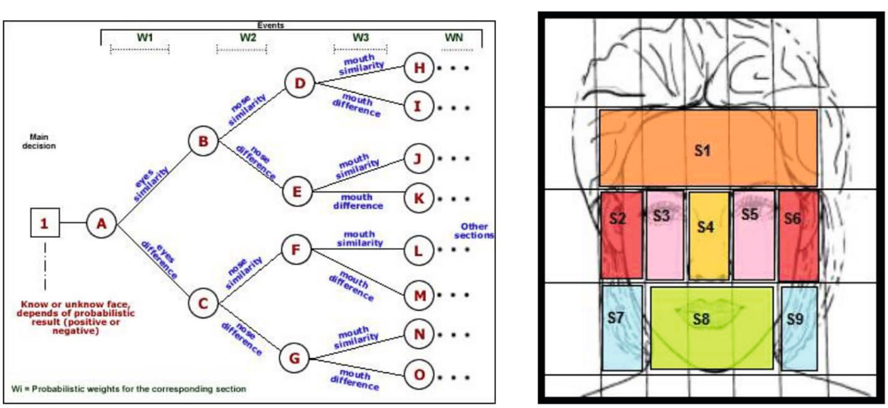



**Fig. 2** Decision tree and Face division in 9 characteristic subsections

accomplish the same characteristics when it is used in a direct way. When this technique is combined with the face subdivisions technique, efficiency is incremented and results are more trustable for faces equality comparison.

### 2.6 Histogram Comparison

An image histogram [8] is a graphical representation of the intensity distribution of an image that quantifies the number of pixels for each intensity value considered. For the face recognition purpose we use the histogram equalization like a way to improve the contrast in an image in order to stretch out the intensity range.

With this first treatment, an image with a more generic tonality was achieved in order to make the recognition process based on the accumulative distribution equation.

$$H^{\hat{}\prime}(i) = \sum_{0 \le j \le i} H(J)$$

The second histogram utilization is a comparison with the saved histograms of known subsections faces in order to be utilized as concept of comparison in the decision tree algorithm for face recognition. To compare two histograms ($H1$ and $H2$), first we have to choose a metric ($d(H1, H2)$) to express how well both histograms match.

For the algorithm, correlation and intersection histograms comparison metrics [8] are utilized.


**Fig. 3** Histogram equalization

## 3 Hybrid Algorithm

### 3.1 Tracking Face

The hybrid algorithm to track a person works as follows:

1. Detection and face cutting: At the beginning of the process, the system detects the person's face using the Haar classifier, once is defined a face is there; it is cut out from the picture and becomes the new interest region.

2. The mass center of the interest region is obtained and stored in a database which will store all coming up mass centers.
3. The stored mass centers are plotted and the Kalman filter is applied to them in order to exclude those items which are far apart.
4. Estimate the position where the person is located.

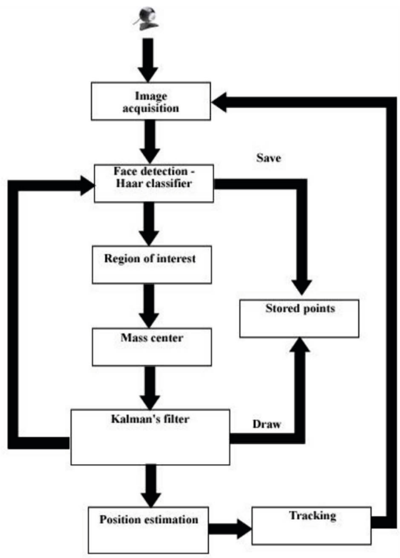



**Fig. 4** Face tracking hybrid algorithm

The hybrid algorithm for face recognition works in three main phases:

1. Detection and cut face: In this first stage, we detect the face of the person using the Haar classifier, once defined that there is a face, this is cut out of the picture.
2. Division of the face: Having defined the image available of the face, this is divided into 9 regions of space.
3. Calculation of similarity: Finally for the algorithm, correlation and intersection histograms comparison metrics [8] are Utilized For Each one of 9 Subsections. Both operations are performed with the detected face and the familiar faces and previously stored. The results of these operations are introduced to the decision tree, by which you get a percentage of similarity of the detected face against a familiar face before.

If the percentage of similarity exceeds a minimum value assigned to the face is considered as known, otherwise, the face belongs to a person not known to the algorithm.

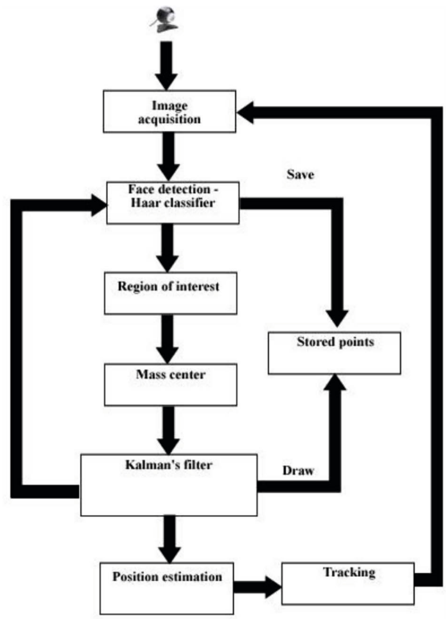



**Fig. 5** Face recognition hybrid algorithm

# 4 Results

## 4.1 Tracking Face

For the implementation of the experiment, we were the participation at least one person to track. When starting up, the person must be placed in front of the camera to detect possible face forward. Once the face is detected locked up his face in a circle and paint the center point of the face and the journey that makes the person as it moves.

Finally, the line graph which would indicate the direction and magnitude at which the person moves. To verify if the plotted course was right, it was proof that a person stands in front of the camera waiting for her face is detected and so get moving. If the program traces the route that matches the person that I do, it is considered that the test was successful.

## 4.2 Face Recognition

This face recognition algorithm was planned to solve the problem of vision test posed "Who is who" for the RoboCup 2012, so for testing and obtaining the same results were a series of tests emulating this specific activity. The evidence that was submitted to the algorithm were to teach four previously

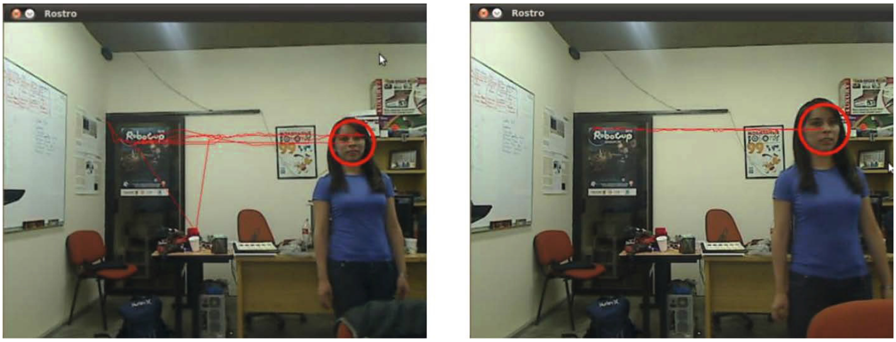



**Fig. 6** Face Tracking and Kalman filter application. The first image on the left is seen to have many disturbances generated by the same variation in the identification of the face, in the second image on the right is applied Kalman and observed substantially improving the correction of the estimated center position of the head

**Table 1** Obtained results for face recognition

| Face Recognition Tests | | |
|---|---|---|
| | Quantity | Average (%) |
| Action decisions | 205 | 100 |
| Correct decisions | 178 | 86.83 |
| Uncorrect decisions | 27 | 13.17 |

(acquaintances) and then start the recognition phase where different people (both known and unknown) were located in front of the camera and then proceeded to run your algorithm procedure to determine whether they were known or unknown to him.

In tests, the system has been exposed to a total of 205 choices of action (situation where you have to decide if a face is known to him or not) of which were obtained the following results:

These tests were performed by taking a static camera images, and showed positive results with a high percentage of correct recognition, as well as efficient processing speed.

## 5 Conclusions

The algorithm was tested in the service robot "Donaxi @ HOME", which participates in the RoboCup since 2009.

The model for face tracking and recognition is proposed in this paper seeks to exploit the ability to use the comparison of characteristics to define whether a face is known or not, based on a current face and the features of a face


**Fig. 7** Face recognition real time application

stored above so we can determine if the face that is always the same. This was determined by face recognition at the beginning and end of follow up.

The combination of different techniques shown here (Haar classifier, face split into subregions, phase correlation, comparison of histograms, decision trees and Kalman filter) have unique capabilities and advantages that may develop in part a recognition and tracking faces, but in combination (to generate the proposed hybrid algorithm) can get a better focus and greater efficiency when performing this task.

The results have shown that one can obtain a high efficiency rate with this algorithm, although the results should be viewed with caution because this algorithm has not yet passed the phase of high-level tests (test the algorithm bases data and robust free faces that are available on the web) so what is the next phase of its growth.

**Acknowledgement.** We want to thank to the Mexican Robotics Federation (FMR) and Robotics and Mechatronics Network of Mexico for the support provided to carry out this work. Also at the National Council of Science and Technology (CONACYT) of Mexico for push forward technological development in the area of service robotics in Mexico.

## References


1. Baek, K., Kim, B., Park, S., Han, Y., Hahn, H.: Multiple Face Tracking Using Kalman Estimator Based Color SSD Algorithm. In: Zhang, S., Jarvis, R.A. (eds.) AI 2005. LNCS (LNAI), vol. 3809, pp. 1229–1232. Springer, Heidelberg (2005)
2. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The FERET Database and Evaluation Procedure for Face-Recognition Algorithms. Image and Vision Computing 16, 295–306 (1998)
3. Computer Vision Research Progress, http://books.google.com.mx/books?id=g9bP-7oBulUC&dq=%22Some+researchers+proposed+different+control+and+noise+models%22&hl=es&source=gbs_navlinks_s
4. Kodratoff, Y., Michalski, R.S.: Machine learning: an artificial intelligence approach, 1st edn., vol. III, pp. 140–146. Morgan Kaufmann (August 1990) ISBN-10: 1558601198
5. Maturana, D., Mery, D., Soto, A.: Face Recognition with Decision Tree-based Local Binary Patterns. Department of Computer Science. Ponticia Universidad Católica de Chile, Chile (2010)
6. Cedeño, J.C.: La cara, sus proporciones estéticas. Clinica Central Cira García, La Habana, Cuba
7. Reddy, B.S., Chatterji, B.N.: An FFT-based technique for translation, rotation, and scale-invariant image registration. IEEE Transactions on Image Processing 5(8), 1266–1271 (1996)
8. OpenCV dev-team, OpenCV v2.3 documentation - Histogram Equalization - Histogram Calculation - Histogram Comparison (August 2011)
9. Quezada, C.V.: Reconocimiento de Rostros Utilizando Análisis de Componentes Principales: Limitaciones del Algoritmo, Universidad Iberoamericana (2005), http://www.bib.uia.mx/tesis/pdf/014620/014620_00.pdf
10. Maturana, D., Mery, D., Soto, Á.: Face Recognition with Decision Tree-Based Local Binary Patterns. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part IV. LNCS, vol. 6495, pp. 618–629. Springer, Heidelberg (2011)

# No-Reference Image Quality Assessment for Iris Biometrics

Valery Starovoitov, Agnieszka Kitlas Golińska,
Anna Predko-Maliszewska, and Maciej Goliński

Department of Medical Informatics, Institute of Computer Science,
University of Białystok, Sosnowa 64, 15-887 Białystok, Poland
valerystar@mail.ru, {akitlas,mgolinski}@ii.uwb.edu.pl, aniapredko@o2.pl

**Summary.** No-reference image quality assessment (NRIQA) methods estimate image quality degradations without any information about the Şperfect-qualityŤ reference image. In this paper, we propose an NRIQA algorithm based on the idea of comparison two blurred variants of the original image to be estimated.

## 1 Introduction

In recent years, there has been an increasing need of accurate and easy-to-use image quality assessment (IQA) algorithms in a variety of applications, including image compression, printing, display, restoration, segmentation, and so on [1]. Most existing IQA methods require full access to an original reference image that is assumed to have perfect quality. Without the reference image, the IQA task becomes very difficult, and almost all existing no-reference IQA measure were designed to detect a certain type of distortion such as blocking and blurring in JPEG [2], blurring [3] and wavelet quantization in JPEG2000 [4]. In practice, these no-reference methods are useful only when the types of distortions between the reference and distorted images are known.

For example, in iris biometrics the scanning process is not very comfortable for user and it would be convenient to do this as fast as possible. This means that an algorithm has to discard blurred iris images in real time. Moreover, we do not have a reference iris image. Thus the appropriate image quality measure must be no-reference, simple and should yield similar results to visual perception of image quality.

Our no-reference image quality measure assumes that the degradation of image quality is caused only by an object being outside of depth of field of the optical registration system. Due to technical limitations, depth of field is often too small to ensure that no part of an image is out-of-focus.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 95–100.
springerlink.com 

To evaluate the image quality we convolve it with two low pass filters of different sizes. If the image is sharp enough then the difference between filtered outputs will be high. We present an algorithm of a simple no-reference measure for blur evaluation and compare it with some other no- and full-reference measures published in the literature.

## 2 Review of Selected IQA Measures

### 2.1 The Daugman's Quality Measure (DA Measure)

Daugman observed that one can assess the quality of the image by analyzing high frequencies of 2-D Fourier transform [5]. In practice however, due to the need to reduce computational complexity, he proposed to use $8 \times 8$ convolution matrix. Final value of DA measure is a normalized sum of squares of the convoluted image.

### 2.2 The Wei's Measure

Wei et al. [6] introduced an algorithm which is a modification of Daugman's measure. $8 \times 8$ convolution matrix was replaced by $5 \times 5$ convolution matrix reducing computational complexity even further.

### 2.3 The Perceptual Blur Metric

Perceptual blur metric is a measure proposed by F. Crete et al. in 2007 [7]. First the image is blurred by a horizontal and a vertical strong low-pass filter. Then we need to compute absolute difference images in both vertical and horizontal directions of original and blurred images. It allows us to study the variation of the neighboring pixels in the original image and we compare it to the variation after blurring. Next step is to compute the sum of absolute difference images and normalize the results to obtain final blur metric value.

### 2.4 The Wan's Measure

Wan et al. [8] proposed an effective method based on "Laplacian of Gaussian operator" for iris image assessment. First one applies Gaussian blur in order to get rid of noise. Subsequently one uses a Laplace operator (second order difference operator) in order to detect edges in the image.

### 2.5 The SSIM Quality Measure

SSIM quality measure is a full-reference measure and it is combine from three components: luminance comparison, contrast comparison and structure comparison [9]:

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{1}$$

where $l$ is a function which depends on means of images $x$ and $y$, $c$ depends on standard deviations of $x$ and $y$ and $s$ depends on cross-correlation of images $x$ and $y$ normalized by removing mean and dividing by standard deviation. $\alpha, \beta, \gamma > 0$ are parameters used to adjust the relative importance between three components.

## 3 The Proposed Measure

We propose our algorithm for image quality assessment, which is designed to detect blur. First we define mask sizes (for filters $k1$ i $k2$). Then we create a blurred image $B1$ by convolution with box filter $k1 \times k1$ and a blurred image $B2$ by convolution with box filter $k2 \times k2$. We obtain a measure of blur $C$ as a sum of squares of difference between blurred images $B1$ and $B2$ divided by width and height of the image:

$$C = \sum_{(i,j)} \frac{(B1(i,j) - B2(i,j))^2}{M \cdot N} \tag{2}$$

where $M$ and $N$ are respectively width and height of the image.

If $C > t$ the test image is considered to be of a low quality (blurred), where $t$ is a chosen threshold. In our examples below we have used parameters $k1 = 9$, $k2 = 21$, $t = 6$.

## 4 Experiments

We have chosen two sets of iris images (6 images in each set). First two images in each set are natural images of iris, the rest images are artificially blurred (Gaussian blur, average blur and motion blur). We sorted images in accordance to image quality. Image 1 is the most sharp, image 6 is the most blurred. In Fig. 1 we presented example of two images from our study: the most sharp image 1 and the most blurred image 6. Size of images was $1000 \times 668$ pixels, 8-bit greyscale. Images used in our study were obtained from the optical system for iris enrollment, which was designed by KBTEM-OMO (Republican Unitary Scientific and Production Enterprise, Minsk, Belarus).

In Fig. 2, Fig. 3, Fig. 4 we present results of our investigation for the first set of images. We applied DA measure, Wei's measure, perceptual blur metric, Wan's measure, SSIM quality measure and our proposed measure (all with Gaussian blur, average blur and motion blur). We obtained similar results for second set of images.

All applied methods distinguish between poor quality images and good quality images, except Wan's measure for Gaussian blurred images. The best

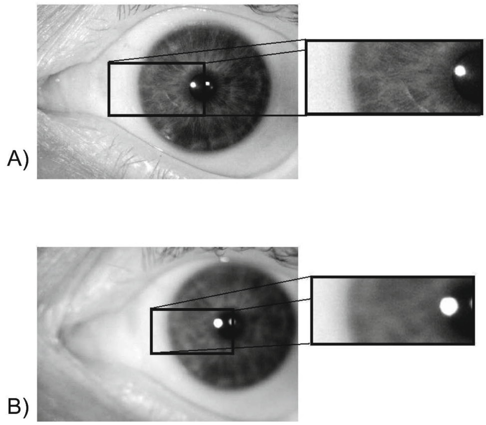



**Fig. 1** Examples of iris images: A) good quality image, B) image with blur

results were obtained from DA measure, perceptual blur metric, proposed measure and full-reference SSIM quality measure. In cases of perceptual blur metric, proposed measure and Wei's measure we can chose threshold equal to approximately 0.8. The DA measure is the most effective measure in our investigating no-reference measures. Also our proposed measure gives similar results for blurred images.

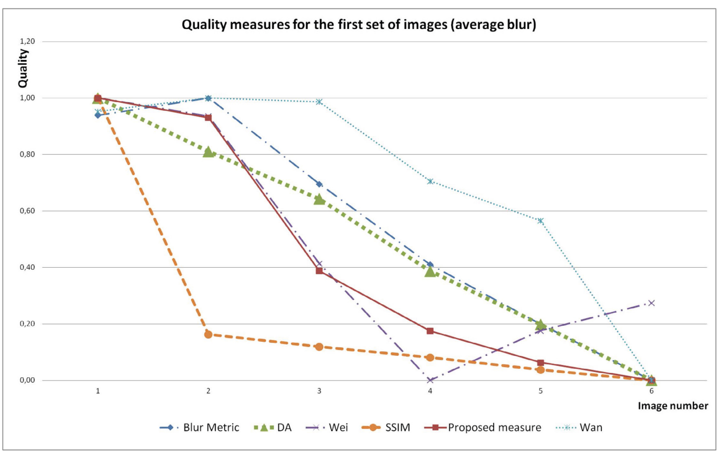



**Fig. 2** Quality measures for the first set of images (average blur). Image 1 is the most sharp, image 6 is the most blurred.

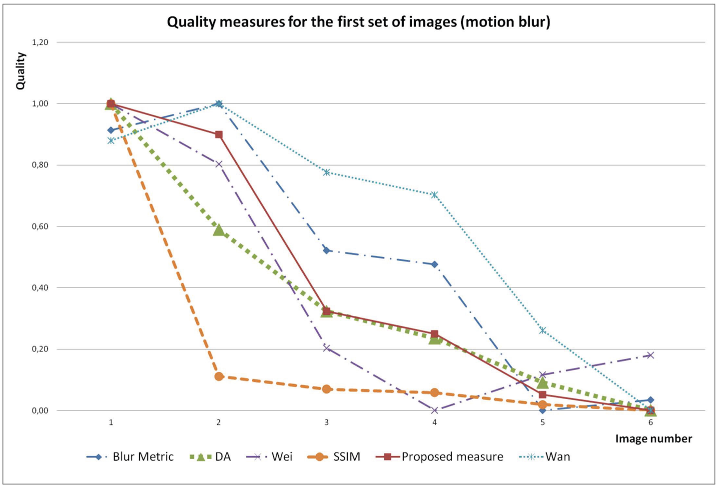



**Fig. 3** Quality measures for the first set of images (motion blur). Image 1 is the most sharp, image 6 is the most blurred.

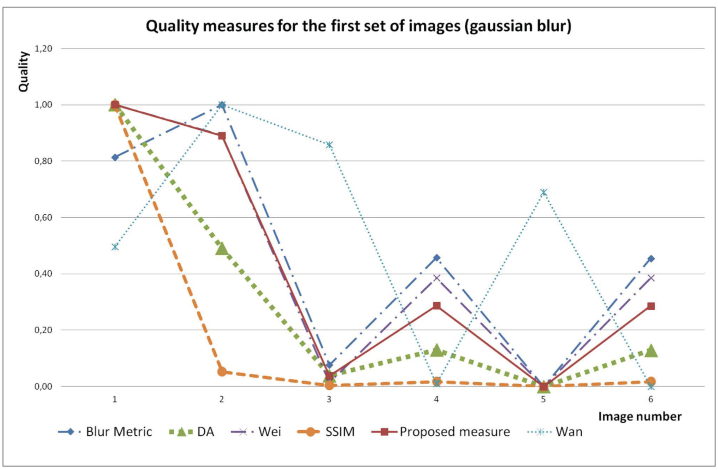



**Fig. 4** Quality measures for the first set of images (Gaussian blur). Image 1 is the most sharp, image 6 is the most blurred.

In case of Gaussian blur all methods, except DA measure and full-reference SSIM quality measure, have difficulties in distinguishing between blurred images.

For the second set of images we obtained similar results for all three types of artificial blurs.

We hope that results obtained from our proposed measure could be improved by better choice of parameters.

## 5 Conclusion

Our work shows that proposed measure can be useful tools in the analysis of iris images. Now we are in a process of testing our measure and choice of parameters for a larger set of images to prove its value.

## References

1. Wang, Z., Bovik, A.C.: Modern image quality assessment. Morgan and Claypool, San Rafael (2006)
2. Wang, Z., Sheikh, H.R., Bovik, A.C.: No-reference perceptual quality assessment of JPEG compressed images. In: Proc. IEEE Int. Conf. Image Process., Rochester, pp. 477–480 (2002)
3. Marziliano, et al.: Perceptual blur and ringing metrics: application to JPEG 2000. Signal Process. Image Commun. 19, 163–172 (2004)
4. Sheikh, R.H., Bovik, A.C., Cormack, L.: No-reference quality assessment using natural scene statistics: JPEG 2000. IEEE Trans. Image Process. 14(11), 1918–1927 (2005)
5. Daugman, J.: High confidence recognition of persons by their iris patterns. In: IEEE 35th International Carnahan Conf. on Security Technology, London, UK, pp. 254–263 (2001)
6. Wei, Z., Tan, T., Sun, Z.: Nonlinear Iris Deformation Correction Based on Gaussian Model. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 780–789. Springer, Heidelberg (2007)
7. Crete, F., Dolmiere, T., Ladret, P., Nicolas, M.: The blur effect: perception and estimation with a new no-reference perceptual blur metric. In: Proc. SPIE, San Jose, vol. 6492 (2007)
8. Wan, J., He, X., Shi, P.: An iris image quality assessment method based on laplacian of gaussian operator. In: IAPR Conf. on Machine Vision Appl., MVA 2007, Tokyo, Japan, pp. 248–251 (2007)
9. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. on Image Process. 13(4), 600–612 (2004)

# Camera Pose Estimation from Sequence of Calibrated Images

Jacek Komorowski[1] and Przemysław Rokita[2]

[1] Maria Curie-Sklodowska University, Institute of Computer Science, Lublin, Poland
jacek.komorowski@gmail.com
[2] Warsaw University of Technology, Institute of Computer Science, Warsaw, Poland
p.rokita@ii.pw.edu.pl

**Summary.** In this paper a method for camera pose estimation from a sequence of images is presented. The method assumes camera is calibrated (intrinsic parameters are known) which allows to decrease a number of required pairs of corresponding points compared to uncalibrated case. Our algorithm can be used as a first stage in a structure from motion stereo reconstruction system.

## 1 Introduction

Motivation for development of the method described in this paper was our prior research on human face reconstruction from a sequence of images from a monocular camera (such as depicted on Fig. 1). Such sequence representing an object moving or rotating in front of a fixed camera can be alternatively thought of as a sequence of images of a static object taken by a moving camera and such perspective is adopted in this paper. Classical multi-view stereo reconstruction algorithms (as surveyed in [9]) assume fully calibrated setup, where both intrinsic and extrinsic camera parameters are known for each frame. Such algorithms cannot be used in our scenario where an object moves or rotates freely in front of the camera. Even if camera intrinsic parameters are known and fixed during the entire sequence, camera extrinsic parameters (rotation matrix and translation vector relating camera reference frame with the world reference frame) are not known. So before a multi-view stereo reconstruction algorithm can be used a prior step to estimate camera pose (extrinsic parameters) for each image in the sequence is required.

Such methods usually work by finding a correspondence between feature points on subsequent images and then recovering camera pose and scene structure using matched features. Human skin has relatively a homogeneous texture and initial experiments showed that methods based on tracking feature points

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 101–109.
springerlink.com 

on subsequent images, such as Kanade-Lucas-Tomasi tracker were not performing as expected. In this paper we describe an alternative approach based on ideas used in modern structure from motion products such as Bundler [6] or Microsoft PhotoSynth. These solutions work by finding geometric relationship (encoded by fundamental matrix) between 2 images of the scene taken from different viewpoints. This is usually done by running a robust parameter estimation method (e.g. RANSAC) combined with 7-point or 8-point fundamental matrix estimation algorithm using putative pairs of corresponding features from 2 images.

However, in contrast to aforementioned solutions, in our method we make an assumption about fixed and known (from a prior calibration stage) camera intrinsic parameters. When camera intrinsic parameters are known, less pairs of corresponding points are required to recover 2-view scene geometry. This should significantly decrease number of iterations needed by RANSAC to estimate model parameters with a given confidence.

Unfortunately currently known algorithms for estimation of relative pose between 2 calibrated cameras from 5 pairs of corresponding points are very complex and implementations are not freely available. E.g. Nister 5-point algorithm [4] requires SVD, partial Gauss-Jordan elimination with pivoting of a system of of polynomial equations of the third degree and finally finding roots of a 10th degree polynomial. Such complexity can potentially lead to significant numerical errors and make such methods inapplicable in practice.

So aim of our work was twofold: first to design a solution for estimation of extrinsic parameters for a sequence of images from a calibrated camera, and second, to verify that C++ implementation of Nister 5-point algorithm is numerically stable.

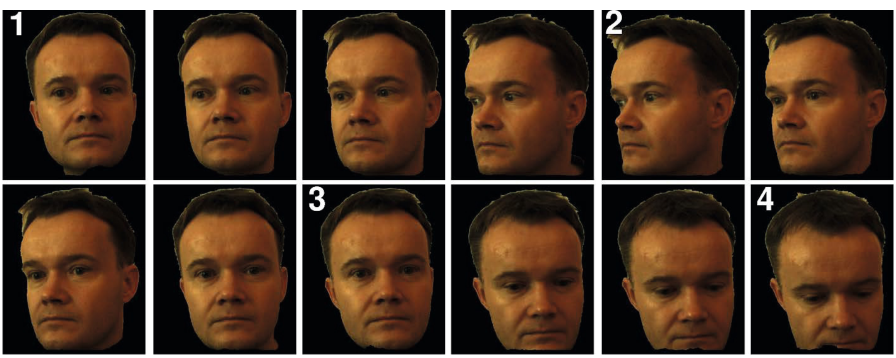



**Fig. 1** Examplary images from an input sequence (every 8th image of 120 frames input sequence). Head is initially positioned in front of the camera (1) and rotates right app. 45°(2), back to the frontal pose (3) and downward app. 20° (4).

## 2 Pose Estimation Method Details

An input to our pose estimation method is a sequence of images from a calibrated, monocular camera, such as depicted on Fig. 1. The the following steps are done:

1. Initial processing: object segmentation from the background and geometric distortions removal. Further processing is done on undistorted and segmented images.
2. Detection of SIFT features on all images in the sequence.
3. Estimation of the relative pose between 2 initial images in the sequence:
   a) Finding pairs of putative matches between SIFT features on both images.
   b) Computation of essential matrix $E_{12}$ relating two images using RANSAC [1] with Nister [4] solution to 5-point relative pose problem. The relative pose (translation vector $T_2$ and rotation matrix $R_2$) is recovered from $E_{12}$ as described in [4].
   c) Construction of an initial 3D model (as a sparse set of 3D points) by metric triangulation of pairs of consistent features (reprojection error and distance between feature descriptors are below thresholds) from two images.
   d) 3D points and camera pose refinement using bundle adjustment method [8] to minimize reprojection error.
4. Iterative estimation of an absolute pose of each subsequent image $I_n$ with respect to 3D model built so far:
   a) Finding putative matches between features on the image $I_n$ and 3D points already in the model.
   b) Computation of an absolute pose (translation vector $T_k$ and rotation matrix $R_k$) of the image $I_k$ with respect to the 3D model. This is done using RANSAC [1] with Finsterwalder 3-point perspective pose estimation algorithm [3].
   c) Guided matching of features from currently processed image $I_k$ and images processed in the previous steps. New 3D points are generated and added to 3D model (and support of existing 3D points is extended) by metric triangulation of matching features.
   d) 3D points and camera pose refinement using bundle adjustment method [8] to minimize reprojection error.

Final results are depicted on Fig. 6, where green dots represent recovered camera poses for each image from an input sequence from Fig. 1.

Additional details on each algorithm step:

*Step 2*

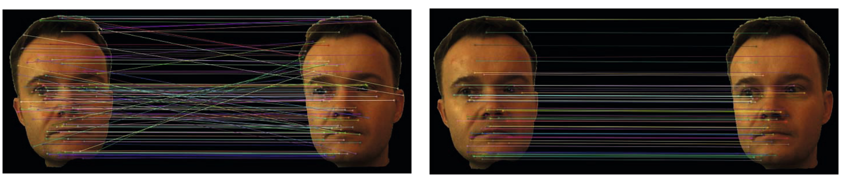
SIFT features [5] are a common choice in modern structure from motion solutions. This is dictated by their invariance to scaling, rotation and, to some extent, lighting variance and small affine image transformations. These properties are important when finding corresponding features on images taken from different viewpoints. At this step SIFT features are found and feature descriptors (represented as vectors from $\mathbb{R}^{128}$) are computed for each image in the sequence.

*Step 3a*

For each keypoint from the first image the closest (in the feature descriptor space) keypoint from the second image is found. Only pairs fulfilling nearest neighbour ratio criterion (that is ratio of a distance to the corresponding keypoint to the distance to the second-closest keypoint on the other image is below given threshold $\Theta = 1.25$) are kept as putative matches. See Fig. 2a.

*Step 3b*

RANSAC [1] robust parameter estimation is used with our implementation of Nister 5-point algorithm [4] to estimate relative pose between 2 cameras from a set of putative point correspondences. Results of this step are: essential matrix $E_{12}$ describing stereo geometry between 2 images from a calibrated camera, rotation matrix $R_2$ and translation vector $T_2$ describing the relative pose of the second image with respect to the first image, consensus set consisting of pairs of matching features consistent with epipolar geometry (see Fig. 2b)



(a) (b)

**Fig. 2** Pairs of matches between 2 images 2a putative matches 2b matches consistent with epipolar geometry encoded by estimated essential matrix $E$

## 3 Experimental Results

Quantitative evaluation of accuracy of the presented method is difficult due to lack of reliable ground truth data. In this paper we only present quantitative examination, using a synthetic data, of one key component of our solution, that is estimation of the relative pose of two calibrated cameras using our implementation of Nister 5-point algorithm.

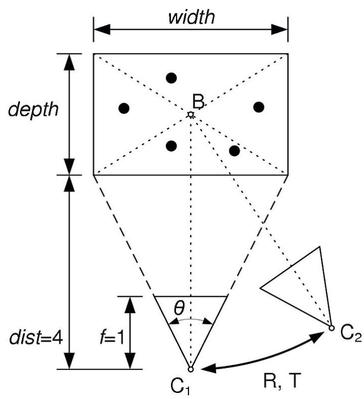



**Fig. 3** Test environment used to generate synthetic data

Configuration used to generate synthetic data for experiments was similar to used in [4] and is depicted on Fig. 3. Random 3D points are generated within an axis-aligned bounding box located in front of the camera. Camera focal local length $f = 1$ and distance between front of the bounding box and camera optical center $dist = 4$. Front of the bounding box subtends $\Theta = 45°$ of visual angle. Width of the bounding box is calculated from the formula: $\frac{width/2}{dist} = \tan \Theta/2 \Rightarrow width = 2dist \tan \Theta/2$ and height is equal to width. Height of the bounding box is equal to width. It's assumed camera center $C_1$ is at the origin of world reference frame. 3D points are rotated by angle $\alpha$ around bounding box center $B$ and the rotation axis is parallel to the world coordinate frame $y$ axis. Suppose 3D point $X$ is rotated around a point $B$ by applying 3x3 rotation matrix $R$. The point coordinates in camera reference frame after rotation are $X' = R(X - B) + B = R(X - (B - R^{-1}B))$. This is equivalent to moving a camera to the new position $C_2$ by applying rotation $R$ and translating camera optical center by a vector $T = B - R^{-1}B$.

Generated 3D points are then projected onto image planes of both cameras and zero mean Gaussian noise is added to projection coordinates. In order to convert noise from pixel units to focal length units we assume x-resolution of a camera image plane is 1296 pixels (which corresponds to high resolution cameras). So $2f \tan \Theta/2 = 1296$ and 1 pixel corresponds to $1/648f \tan \Theta/2$ of focal length units.

The aim of the experiment was to study accuracy of camera rotation and translation estimation based on noisy projection coordinates using our implementation of Nister 5-point algorithm and compare it to much simpler 7-point algorithm. In each experiment $N = 10000$ trials were performed. In each trial camera rotation $\hat{R}$ and translation $\hat{T}$ were estimated from noisy projection coordinates and compared with ground truth rotation $R$ and translation $T$. Experiments were performed for 3 different configurations: almost-planar (depth to width ratio of the bounding box in which 3D points were generated = 0.01), semi-planar (depth to width ratio = 0.1), general (depth to width ratio = 1.0).

As metric reconstruction based on images from 2 calibrated cameras (only intrinsic parameters are known) is possible only up to a scale factor we cannot directly compare true translation vector $T$ and estimated translation vector $\hat{T}$. Only angular component of translation error, that is an angle between true translation vector $T$ and estimated translation vector $\hat{T}$, is calculated using the formula:

$$T_{err} = \cos^{-1}\left(\frac{\hat{T}_i \cdot T}{|\hat{T}_i||T|}\right)$$

Rotation error $R_{err}$ is measured as the rotation angle needed to align ground truth rotation matrix $R$ and estimated matrix $\hat{R}$.

$$R_{err} = \cos^{-1}\frac{\mathrm{Tr}\left(\Delta R\right) - 1}{2},$$

where $\Delta R = R^{-1}\hat{R}$ is the rotation matrix that aligns estimated rotation $\hat{R}$ with the true solution $R$ and $\mathrm{Tr}(\Delta R)$ is a trace of $\Delta R$.

Experiment results are depicted on Fig. 4 and 5. Fig. 4 presents median, first and third quartile of rotation matrix estimation errors $R_{err}$ for $N = 1000$ trials. It can be seen that for all tests scenarios: almost-planar (left column),

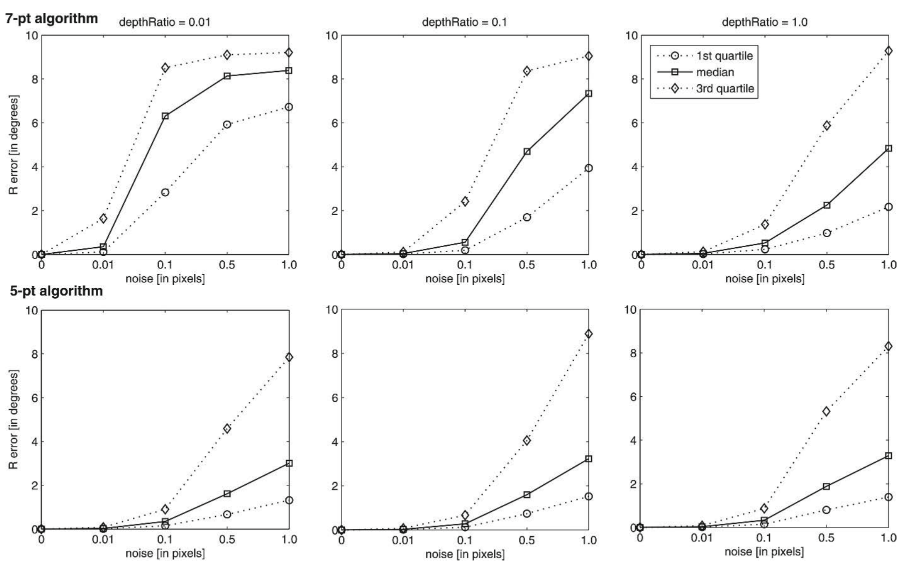



**Fig. 4** Rotation matrix estimation error $R_{err}$. Top row: 7-point algorithm, bottom row: 5-point algorithm. Columns from left to right correspond to almost-planar, semi-planar and general configurations.

semi-planar (middle column) and general (right column) rotation error $R_{err}$ median is lower for 5-point algorithm than for 7-point algorithm. Additionally 5-point algorithm does not suffer from planar degeneracy, it performs equally well in all 3 configurations (almost planar, semi planar and general). Performance of 7-point algorithm deteriorates significantly when 3D point configuration becomes more planar, which is in line with theoretical results.

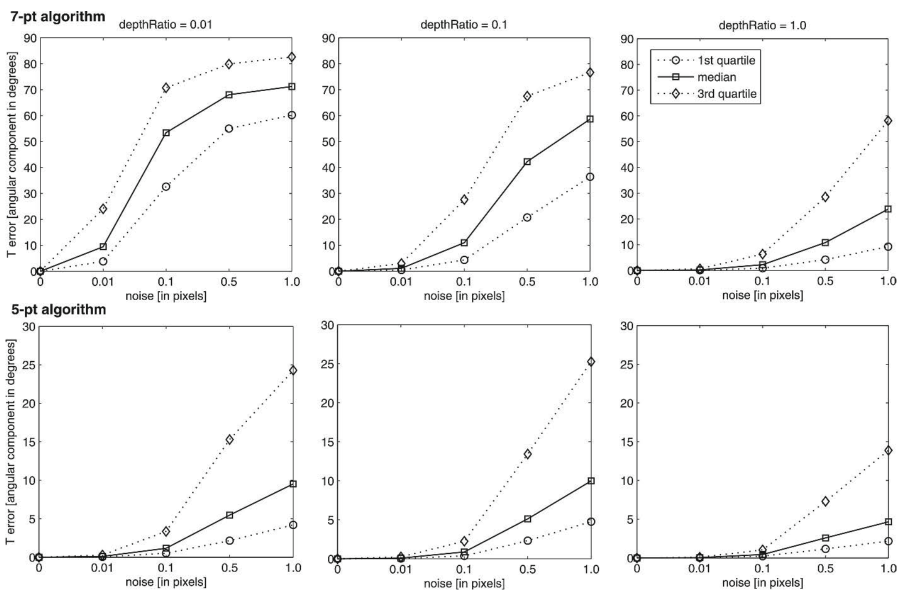



**Fig. 5** Translation vector estimation error $T_{err}$ (angular component only). Top row: 7-point algorithm, bottom row: 5-point algorithm. Columns from left to right correspond to almost-planar, semi-planar and general configurations.

Results for translation error $T_{err}$ are presented on Fig. 5. Again 5-point algorithm outperforms 7-point algorithm in all 3 configurations. It can be noted that 5-point algorithm translation error increases when 3D point configuration becomes more planar but still is significantly lower than 5-point algorithm error.

*Experiments on real world data*

Camera pose reconstruction results on a real world data was verified using a number of input sequences containing images of a head rotating in front of the camera. Camera pose estimation results on exemplary input sequence depicted on Fig. 1 are shown on Fig. 6. Results are as expected: head is initially placed in front of the camera and it corresponds to camera placed at position 1, then as head rotates rightward camera moves along a circular trajectory up to position 2, then head rotates leftward, back to a frontal pose, and camera moves backward along a circular trajectory to position 3, finally head is rotates downward which corresponds to camera moving up to position 4. But due to lack of a reliable ground truth data it was not possible to evaluate these results quantitatively.

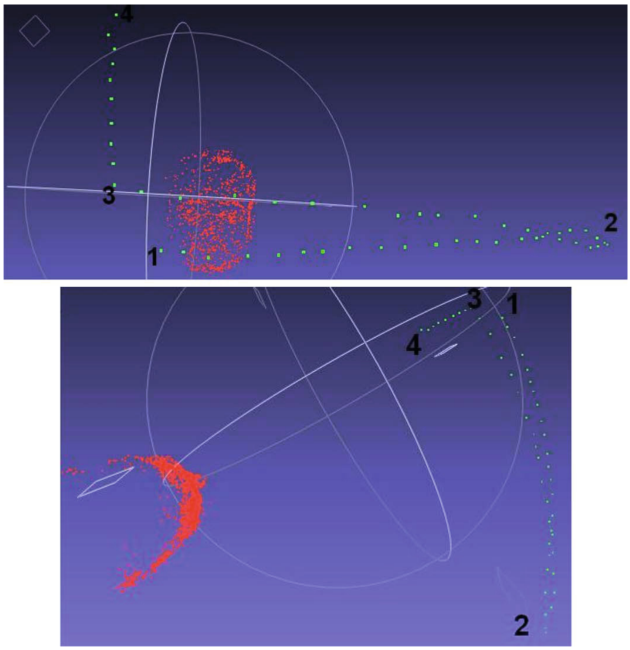



**Fig. 6** Estimated camera positions (green dots) and 3D points (red dots) recovered from sequence depicted on Fig. 1. Head is initially positioned in front of the camera (1) and rotates right app. 45°(2), back to the frontal pose (3) and downward app. 20° (4).

## 4 Conclusions and Future Work

Making an assumption about camera calibration (known intrinsic parameters) allows to find a relative pose between 2 cameras using only 5 pairs of corresponding points instead of 7 pairs needed in uncalibrated setting. This can make a difference when processing images of a low textured objects where a number of reliable matches between features on 2 images is low. It was also quantitatively verified using synthetic data that Nister solution to 5-point relative pose problem, despite its complexity, is numerically much more stable than 7-point algorithm. Additionally performance of 5-point algorithm doesn't deteriorate when 3D points configuration becomes more planar.

In the future it's planned to use the presented method as a first stage in a dense stereo reconstruction system. After camera pose is estimated for each image in the sequence some multi-view stereo reconstruction method will be used to generate a dense point cloud representing an object.

## References


1. Fischler, M., Bolles, R.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM (1981)
2. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2004)
3. Haralick, R., Lee, C., Ottenberg, K., Nolle, M.: Review and analysis of solutions of the three point perspective pose estimation problem. International Journal of Computer Vision (1994)
4. Nister, D.: An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence (2004)
5. Lowe, D.: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision (1999)
6. Snavely, N., et al.: Modeling the World from Internet Photo Collections. International Journal of Computer Vision (2007)
7. Tomasi, C., Kanade, T.: Detection and Tracking of Point Features. Carnegie Mellon University Technical Report (1991)
8. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: Bundle Adjustment Ů A Modern Synthesis. In: Proceedings of the IWVA (1999)
9. Seitz, S., et al.: A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In: CVPR 2006, vol. 1 (2006)

# Significance of Entropy Coding in Contemporary Hybrid Video Encoders

Damian Karwowski

Poznan University of Technology, Chair of Multimedia Telecomunications and Microelectronics
dkarwow@multimedia.edu.pl

**Summary.** The paper presents results on significance of entropy coding used in contemporary hybrid video encoders. Experiments were done with reference to the international video compression standard MPEG-4 AVC/H.264 and the state-of-the-art Context-based Adaptive Binary Arithmetic Coding (CABAC) entropy encoder. Research were done for two different coding scenarios. Experimental results revealed, that application the adaptive entropy encoder within new generation video encoder reduces the size of encoded bitstream 1.5 - 1.8 times, depending on research scenario.

## 1 Introduction

Numerous video compression techniques were worked out in order to efficiently represent video data [1, 2, 3, 5]. Greatest popularity were gained by hybrid coding schemes that found practical applications in many international and commercial video compression standards [1, 2, 5]. In this technology, moving pictures are represented in a compact form by the use of intra-frame prediction, inter-frame prediction, transform coding and entropy coding of residual data. Three first coding techniques are used to de-correlate video signal in both the spatial and the time domains. Their significance and influence on compression performance of video encoder were tested well and described in the literature [1, 2, 3]. It is commonly said that the 'power' of hybrid video encoders just comes from prediction and transform coding mechanisms, and entropy coding is only used to additionally reduce redundancy in resulted residual data. Nevertheless, there are no proper experimental results in the literature that would present real 'power' of entropy coding in the framework of contemporary hybrid video encoders.

This paper answers this question. The answer strongly depends on the type of video encoder and the type of entropy coding used. Successive generations of hybrid video encoders use different mechanisms of forming the residual data and exploit entropy coding techniques of different coding efficiency [1, 2, 3]. An interesting issue is to explore the problem with reference to hybrid video encoder of the new generation.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 111–117.
springerlink.com 

This paper presents research results obtained in the context of the newest international video compression standard MPEG-4 AVC/H.264 [3, 5] and the state-of-the-art Context-based Adaptive Binary Arithmetic Coding (CABAC) entropy encoder [4, 7]. Research were done according to two different coding scenarios.

## 2 MPEG-4 AVC Video Encoder - Entropy Coding

MPEG-4 AVC/H.264 video coding standard has been well presented in the literature, detailed descriptions of the codec can be found in [3, 5]. Therefore, only short presentation of the standard is introduced here from the point of view of entropy coding. In order to form residual data, MPEG-4 AVC encoder exploits sophisticated techniques of intra- and inter-frame prediction of image samples with numerous directional prediction modes (intra prediction) and prediction of image blocks of variable size (inter prediction). In order to additionally de-correlate resulted prediction error samples transform coding is used with DCT-like transformation and transform coefficients quantization. All these operations generate three data streams (quantized transform coefficients, motion data and control data) that are finally encoded with entropy encoder. In the MPEG-4 AVC standard two different entropy coding techniques were defined. These are: less computationally complex but also less efficient Context Adaptive Variable Length Coding (CAVLC) that is used together with Exp-Golomb coding, and more computationally complex but more efficient Context-based Adaptive Binary Arithmetic Coding (CABAC) [4, 7]. Due to the fact that application of CABAC entropy encoder dominates in high definition multimedia systems it will be the scope of further interest in the paper. Figure 1 presents block diagram of CABAC encoder.

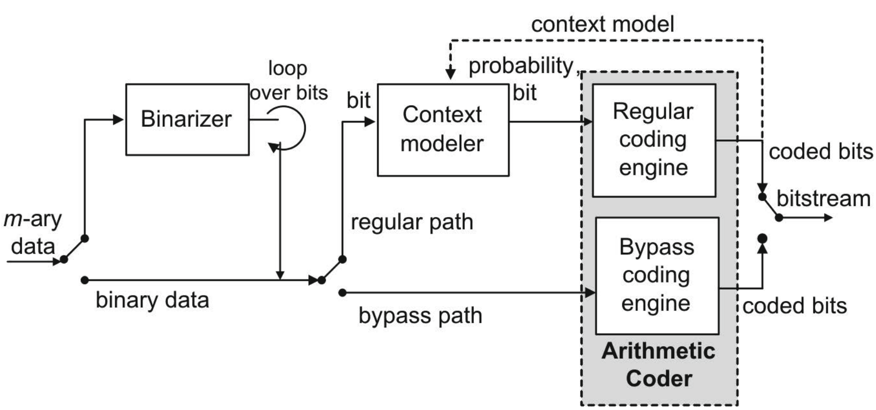



**Fig. 1** Block diagram of CABAC entropy encoder

CABAC reduces redundancy of residual data by the use of well known in data compression arithmetic coding. Arithmetic encoder encodes input symbols with respect to the probability of their occurrence in data stream. The way of data statistics estimation strongly influences compression performance of the whole entropy encoder. Therefore, in order to track statistics of residual data efficiently, CABAC uses advanced mechanism of data probability estimation. First of all, CABAC uses the core of binary arithmetic encoder (instead of m-ary arithmetic encoder core) in order to speed-up computations and limit the problem of context dilution that decreases efficiency of entropy coding. Therefore all non-binary valued data are converted into string of binary symbols in the binarizer block. The size of this string strongly affects compression performance as well as computational complexity of entropy encoder. From that reason, CABAC uses several different binarization schemes that capture features of encoded data, in order to keep the size of binary symbols string as short as possible. Secondly, statistics of symbols in the binarized word are estimated by the use the total number of 460 statistical models. Statistical models were grouped into sub-sets and individual sub-set was assigned to individual syntax elements defined in MPEG-4 AVC. This makes the first level of algorithm adaptation to the current signal statistics. In the second level of algorithm adaptation, the proper statistical model from the sub-set is chosen based on context of encoded data - it means values of a given data symbol in neighboring image blocks.

Such an efficient approach is used for majority types of data encoded in MPEG-4 AVC when using CABAC. For only small sub-set of data symbols that are characterized with uniform probability distribution less computationally complex bypass coding path is used with no data probability estimation.

## 3 Research Problem

Many experimental results revealed that CABAC algorithm outperforms other standardized entropy coding techniques used in hybrid video encoders [4]. The question is in what extent application of efficient CABAC algorithm reduces the size of data in video encoder. In order to answer this question experiments were done for two different coding scenarios performed in the framework of MPEG-4 AVC video encoder.

## 4 Methodology of Experiments

Solution of the problem that was stated above depends on assumed methodology of experiments and type of video encoder used for tests. Therefore, research scenario adopted by author is presented in details here. In order to explore the problem within contemporary hybrid video encoder, experiments were done in the framework of new MPEG-4 AVC video compression standard and the state-of-the-art CABAC entropy encoder used in MPEG-4 AVC. Reference software JM 18.2 of MPEG-4 AVC video codec (encoder and decoder) was used for research [6]. The software was modified by author to be able to work not only with CABAC entropy encoder (original coding path, no software changes) but also without entropy coding stage (modified coding path, omitting CABAC entropy coding). In general, in original coding path residual data is encoded with variable number of bits using CABAC entropy coding that exploits statistics of coded data. In the modified coding path it is assumed that residual data is encoded with no respect to data statistics using fixed number of bits, so it means encoding data without entropy coding. The number of bits in a fixed-length word must allow to represent total dynamic range of residual data that is a result of prediction- and transform-based coding in the video encoder. In this work, the dynamic range of data is calculated for each frame independently based on results of video encoding for each video frame. Based on this information, necessary number of bits in a fixed-length word is calculated for each video frame for individual syntax elements encoded in MPEG-4 AVC. Due to the fact, that motion data and quantized transform coefficients make majority part of encoded bitstream in MPEG-4 AVC, the dynamic range calculation operations are performed for these data types only. Other syntax elements (that make only fractional part of encoded bitstream) are encoded in unmodified manner (as in original MPEG-4 AVC encoder). Nevertheless, such a simplification does not influence experimental results.

In hybrid video encoder, after the quantization some of the transform coefficients have zero values. These transform coefficients are efficiently encoded in original MPEG-4 AVC encoder by marking the position of non-zero valued transform coefficients in a transform block. In a modified MPEG-4 AVC encoder (that works with no entropy coding) two methods of coding the quantized transform coefficients were implemented. In the first method, all transform coefficients (both the zero and non-zero valued) are encoded using pre-calculated fixed number of bits. In the second method, only non-zero valued transform coefficients are encoded using pre-calculated fixed number of bits and zero valued coefficients are encoded in an efficient manner, like in original MPEG-4 AVC encoder.

All experiments were done with three versions of the software: 1) original MPEG-4 AVC with CABAC, 2) modified MPEG-4 AVC with no entropy encoder and the first (less efficient) method of coding the transform coefficients, 3) modified MPEG-4 AVC with no entropy encoder and the second (more efficient) method of coding the transform coefficients. Coding efficiency of modified MPEG-4 AVC encoders was explored and compared against efficiency of the original MPEG-4 AVC video encoder. Experiments were done using the following scenario:

- Four Full HD test video sequences were used: *BQTerrace, Cactus, Kimono1, ParkScene* (spatial resolution: 1920x1080, temporal resolution: 24, 50, 60 Hz);
- IBBPBBP ... group of pictures (GOP) structure was used;
- Experiments were done for different quantization parameter (QP) values: QP=22, 27, 32, 37. This parameter is responsible for the quality of decoded video, assumed range corresponds to quality of reconstructed video frames from excellent to poor.
- In order to have the same control decisions (the same coding modes chosen for individual image blocks) in both the original and the modified encoders, rate-distortion optimization mechanism was switched off.

## 5 Experimental Results

Experimental results showed significantly better performance of original MPEG-4 AVC with CABAC, relative to modified video encoders. Results for two versions of modified MPEG-4 AVC encoders were presented in Figure 2 and Figure 3 respectively.

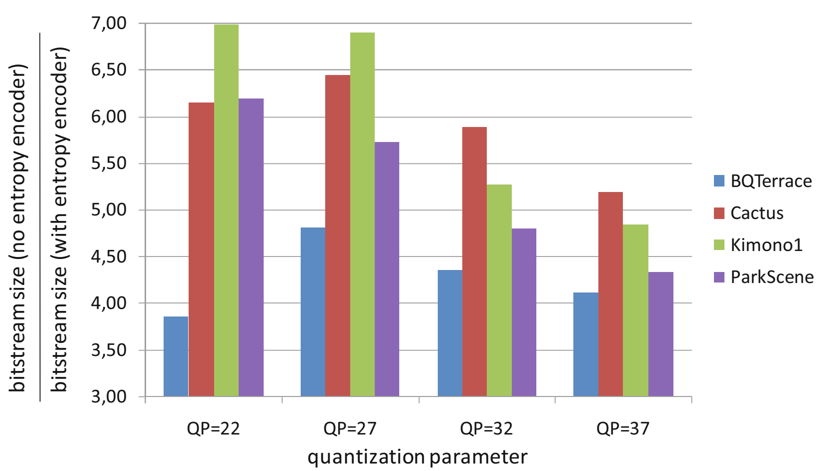



**Fig. 2** Increase of the bitstream size generated by modified MPEG-4 AVC (with first less efficient method of transform coefficients coding) relative to size of bitstream generated by original MPEG-4 AVC

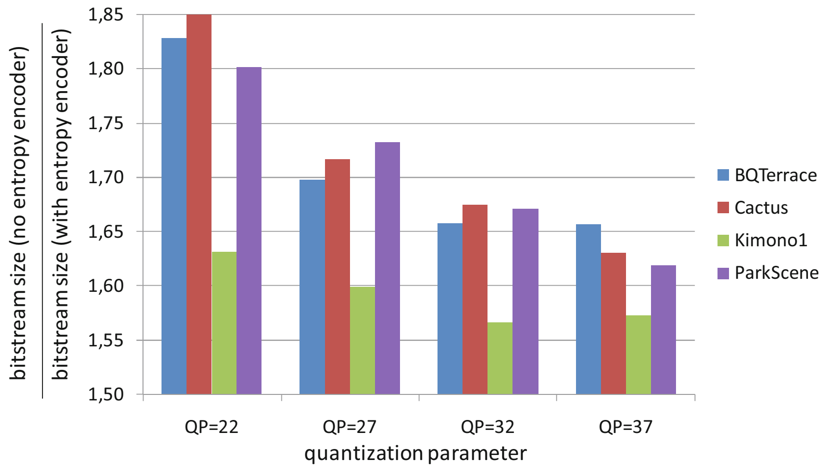



**Fig. 3** Increase of the bitstream size generated by modified MPEG-4 AVC (with second more efficient method of transform coefficients coding) relative to size of bitstream generated by original MPEG-4 AVC

Figure 3 presents results for the version of modified MPEG-4 AVC encoder with less efficient method of transform coefficients coding (both the zero and non-zero valued coefficients are encoded using fixed-length coding). Due to direct encoding of zero valued coefficients using fixed-length words the difference in compression performance is really huge relative to original MPEG-4 AVC with CABAC (even 7-times increase of bitstream size). Mechanism of efficient encoding the zero-valued coefficients is actual not a part of entropy encoder. Therefore, Figure 3 additionally presents results for the version of modified MPEG-4 AVC with more efficient method of transform coefficients coding (only non-zero coefficients are encoded using fixed-length coding, zero coefficients are encoded efficiently like in original MPEG-4 AVC). The results shows that application the CABAC in MPEG-4 AVC reduces the size of bitstream by a factor of 1.5 - 1.8.

The gap in compression performance between original and modified video encoders strongly depends on value of quantization parameter QP. Higher QP value leads to smaller dynamic range of resulted quantized transform coefficients that can be encoded using smaller number of bits in a fixed-length word in modified MPEG-4 AVC. Opposite situation takes place for smaller values of QP. Therefore, the differences in bitstream sizes between original and modified encoders is smaller for higher QPs and higher for smaller QPs.

Different results were obtained for different test video sequences. Test sequences are characterized by different complexity of textures - dynamic range of residual data is different that affects the performance of both the CABAC encoder and the fixed-length coding.

## 6 Conclusions

Results on significance of adaptive entropy coding in hybrid video compression were presented in the paper. Results revealed that adaptive entropy coding techniques are of great importance and significantly reduces the size of data in contemporary hybrid video encoders. It was experimentally proved that the state-of-the-art CABAC entropy encoder reduces the size of data by a factor of 1.5 - 1.8 when activated in the framework of MPEG-4 AVC.

From the compression performance point of view, it is clear that the two parts of video encoder: residual data forming part and entropy coding part should not be analyzed independently. These two parts of encoder works efficiently only when using together. Entropy coding works efficiently for de-correlated data only and data de-correlation makes sense only when realizing in combination with entropy coding stage. It means that the 'power' of video encoder is a result of conjugation the residual data forming techniques and entropy coding methods.

## References

1. Domański, M.: Obraz cyfrowy, Wydawnictwa Komunikacji i Lacznosci (2011)
2. Woods, J.W.: Multidimensional Signal, Image, and Video Processing and Coding. Academic Press (2012)
3. Richardson, I.E.G.: The H.264 Advanced Video Compression Standard, 2nd edn. Wiley (2010)
4. Marpe, D., Schwarz, H., Wiegand, T.: Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard. IEEE Transactions on Circuits and Systems for Video Technology 13(7), 620–636 (2003)
5. Special issue on H.264/AVC video coding standard. IEEE Trans. Circuits and Systems for Video Technology 13 (July 2003)
6. H.264/AVC Software Coordination Site, http://iphome.hhi.de/suehring/tml/
7. Tian, X., Le, T.M., Lian, Y.: Entropy Coders of the H.264/AVC Standard. Springer (2011)

# Comparative Analysis of Image Fusion Performance Evaluation Methods for the Real-Time Environment Monitoring System

Maciej Wielgus[1] and Barbara Putz[2]

[1] Institute of Micromechanics and Photonics, Sw. Andrzeja Boboli 8 St., 02-525 Warsaw, Poland
maciek.wielgus@gmail.com

[2] Institute of Automatic Control and Robotics, Sw. Andrzeja Boboli 8 St., 02-525 Warsaw, Poland
bputz@mchtr.pw.edu.pl

**Summary.** We discuss and compare several objective measures used for image fusion algorithm performance evaluation. Subjective assessments are given as well. Many of the considered evaluation methods originate from prior literature, we also introduce measure based on Jensen-Shannon divergence and a simple gradient-based measure, particularly well fitted for the real time fusion evaluation issue. Along with several well known fusion methods we put under tests recently developed, promising algorithm based on the fast and adaptive bidimensional empirical mode decomposition.

## 1 Introduction

Image fusion consists of merging two or multiple images so that the valuable information from each of the inputs is represented in the output. Typical image fusion problem may involve input images of the same scene acquired by different sensors (multimodal image fusion, e.g., in medical diagnostic problems) or camera images obtained for several different depths of field (multifocus image fusion). This paper is illustrated with multisensor image fusion results with inputs originating from infrared (IR) and visual (VIS) cameras. The data were acquired for the project aimed at the development of the real-time video fusion environment monitoring system [1, 2].

For some applications subjective evaluation of fusion quality performed by human personnel is satisfactory. However, there is an obvious demand for objective measures of algorithm performance whenever comparability and reproducibility of assessments is needed or large amount of data demand automatic evaluation. The latter is the case, e.g., with the real-time video fusion, where fast and efficient automatic fusion algorithm performance evaluation is desirable for quality maintenance reasons. Development of the evaluation method adequate for such an application is within the scope of this paper.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 119–126.
springerlink.com 

Defining an useful, objective and universal measure for image fusion algorithms performance evaluation is a challenging task. Main reason for such a situation is because it is most often ambiguous, which form should the perfect fusion result take. Typical measures of image quality, e.g., mean square error (MSE) or image quality index Q to quantify the structural distortion between two images [3] involve comparison with a reference image. In image denoising, for instance, algorithm performance could be measured for synthetic data by comparing algorithm results with the original, unspoiled image as the reference, representing the desired result of the perfect denoising. A single case in which similar approach to testing fusion results could be applied is when inputs consist of sharp images blurred selectively and the unblurred image serves as the refference. This simulates the multifocus fusion problem and such tests were reported in [4]. However, we confine to the multisensor IR-VIS fusion issue and no perfect reference approach, consequently.

## 2 Fusion Algorithms

Dozens of algorithms were proposed for image fusion purposes. Many of the popular methods involve certain decomposition of the input images on discrete set of scales and subsequent synthesis of the corresponding decomposition elements, but fusion algorithms variety is not limited to such an approach [5]. The experimental part of this paper is illustrated with fusion results and performance evaluations of several common algorithms:

- simple mean value of two input images (MEAN),
- discrete wavelet transform, in this case based on the DBSS(2,2) wavelet (DWT),
- shift invariant wavelet transform [6], in this case based on Haar wavelet (SIDWT),
- Laplacian pyramid (LAPL),
- morphological pyramid fusion (MORPH),
- principal component analysis based fusion (PCA).

Into our considerations we also include recently developed [7, 8] fusion algorithm based on the fast and adaptive bidimensional empirical mode decomposition [9] (FABEMD, here denoted as FAB), for which quantitative evaluation of results quality was not yet reported. The method is particularly interesting in the context of the real-time fusion, as it combines advantages of fusion based on empirical mode decomposition approach [10] with significantly improved time performance.

## 3 Evaluation Methods

In the absence of a reference image representing the desired fusion result, fusion quality measures typically aim to estimate the fraction of meaningful input information retained in the resulting image. Different definitions of information and its importance lead to various quality measures, possessing significantly different properties. To enlist all reported measures is beyond the scope of this paper, but we aim at presenting a representative survey of the most acknowledged methods. Image fusion performance evaluation algorithms can be roughly divided into groups shown in Fig. 1. Only measures that do not demand a reference image are considered further.

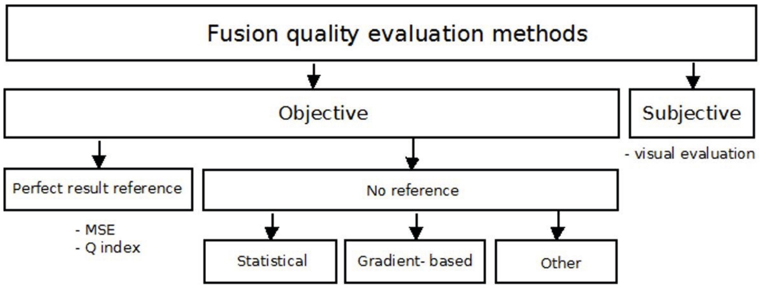



**Fig. 1** Image quality measures classification

## 3.1 Statistical Measures

Statistical measures treat images as random variables subjected to a certain probability distributions, value in a given pixel being a single realization of the corresponding random variable. Consequently, probability theory concepts connected to the distance between random variables and probability distributions can be applied to the issue of images similarity evaluation.

Relation between the image and the corresponding random variable $X_i, i \in \{1, 2\}$ is established as follows. Marginal probability distribution function $p_i(x_i)$ of a random variable $X_i$ is approximated by the $i$-th image histogram normalized by the number of pixels. Similarly, the joint probability distribution $p_{12}(x_1, x_2)$ can be calculated. With so-defined random variables $X_i$ and their estimated distributions $p_i(x_i)$, probabilistic mutual information (MI), defined in Eq. 1, is employed.

$$I(X_1, X_2) = \sum_{x_1, x_2} p_{12}(x_1, x_2) \log \left( \frac{p_{12}(x_1, x_2)}{p_1(x_1) p_2(x_2)} \right) \tag{1}$$

The sum in Eq. 1 is taken over all discrete values of intensity represented in the histograms. MI can be understood as a measure of random variables $X_1$ and $X_2$ mutual dependance. MI-based measure of fusion performance $\mathbf{M_F}$ proposed in [11] is

$$M_F = I(X_1, F) + I(X_2, F) \tag{2}$$

where $F$ denotes the fused image variable of distribution $p_F$. MI is related to the Shannon entropy $H(X_i)$ as

$$I(X_1, X_2) = H(X_1) + H(X_2) - H(X_1, X_2) \tag{3}$$

where Shannon entropy of $X_i$ is defined as

$$H(X_i) = -\sum_{x_i} p_i(x_i) \log p_i(x_i) \tag{4}$$

and the joint entropy of $X_1, X_2$ is

$$H(X_1, X_2) = -\sum_{x_1,x_2} p_{12}(x_1, x_2) \log p_{12}(x_1, x_2) \tag{5}$$

Shannon entropy $\mathbf{H}$ itself may serve as a simple measure of amount of information present in the image, indirectly indicating the fusion quality with no information of source images. Different entropy definitions lead to alternative mutual information prescriptions. For fusion quality assessment Tsallis entropy [12] and Renyi entropy [13] were proposed. In particular Tsallis entropy is defined as

$$H^{\alpha}(X_i) = (1-\alpha)^{-1} \left( \sum_{x_i} p_i(x_i)^{\alpha} - 1 \right) \tag{6}$$

with a parameter $\alpha$. It can be understood as a generalization of the Shannon entropy. In fact, using L'Hospital's rule Eq. 4 is obtained from Eq. 6 for $\alpha \to 1$. Tsallis entropy leads to the following definition of the mutual information

$$I^{\alpha}(X_1, X_2) = \frac{1}{1-\alpha} \left( 1 - \sum_{x_1,x_2} \frac{p_{12}(x_1, x_2)^{\alpha}}{(p_1(x_1) p_2(x_2))^{\alpha-1}} \right) \tag{7}$$

and the fusion performance measure $\mathbf{M}_{\mathbf{F}}^{\alpha}$ defined similarly as in Eq. 2. MI for the Renyi entropy differs from Eq. 7 only by a constant value of $(1-\alpha)^{-1}$. Both [12] and [13] report improved measure performance for $\alpha \approx 2$, therefore we use $\alpha = 2$ in the experiments.

A different probabilistic concept connected to the notion of objects similarity is the Kullback-Leibler divergence, which measures pseudo-distance between probability distributions.

$$D_{KL}(p_1||p_2) = -\sum_{x_i} p_1(x_i) \log \frac{p_1(x_i)}{p_2(x_i)} \tag{8}$$

The Shannon MI (Eq. 1) itself may be viewed as a Kullback-Leibler divergence of the distributions $p_{12}(x_1, x_2)$ and $p_1(x_1)p_2(x_2)$. Based on $D_{KL}$, so-called Jensen-Shannon divergence can be defined:

$$D_{JS}(p_1||p_2) = 0.5 \left[ D_{KL}\left(p_1||m\right) + D_{KL}\left(p_2||m\right) \right] \tag{9}$$

for $m = 0.5(p_1 + p_2)$. It possesses properties valuable for evaluation purposes, such as symmetricity and being confined to a $[0, 1]$ interval (both unlike $D_{KL}$). We propose a quality measure $\mathbf{M_F^{JS}}$ based on $D_{JS}$

$$M_F^{JS} = 1 - D_{JS}(p_1||p_F) - D_{JS}(p_2||p_F) \tag{10}$$

Statistical measures have solid basis in probability and information theory. However, they show disturbing properties such as favoring simple averaging, while in most cases it is clearly outperformed by any of the more sophisticated methods. Trivial example is when certain object is visible only in one of the inputs. Averaging reduces its visibility twice, while other algorithms are capable of retaining the object unspoiled.

### 3.2 Gradient-Based Measures

Gradient-based measures assume that the perceptually important information that should be transferred to the fused image is rather an edge information than the overall statistical similarity. Therefore similarity of gradients is to be considered. This assumption fits well with our knowledge of the human vision mechanisms. The method based on calculations of similarity between gradients magnitudes and orientations was proposed in [14]. In this approach local gradient mismatch indicators are weighted according to the local gradient magnitude, so errors in the edge pixels have more influence on the proposed measure $\mathbf{Q^{AB/F}}$ value than errors in the homogeneous regions. Exact value of $Q^{AB/F}$ depends on seven independent constant parameters. Industrial application of $Q^{AB/F}$ index was presented in [15].

We propose another gradient-based measure, simply weighted gradient error **WGE**. Not only it does not involve any arbitrary parameters, it is also much less computationally expensive than $Q^{AB/F}$. Denoting $E_{ij} = S_j(F - X_i)S_j(X_i)$, where $X_i$ are the input images and $S_j$ is the Sobel operator in $j$-th spatial direction, $i, j \in \{1, 2\}$ we have

$$WGE = \frac{\left(\sum_{i,j} E_{ij}^2\right)^{1/2}}{4NMD} \tag{11}$$

$N, M$ are the numbers of image rows and columns and $D = \max(F) - \min(F)$.

### 3.3 Other Objective Measures

In [4] two measures based on Q image quality index were introduced. In this approach Q index of similarity between $X_i$ and $F$ is first calculated locally for each pixel of each input image. Then the local saliency of images is calculated, assumed to be proportional to the local image variance. For each pixel$(n, m)$, $Q_i(n, m)$ values for both inputs are averaged with the saliency-based weights to obtain $Q(n, m)$. Finally, $\mathbf{Q_W}$ index is obtained as a weighted sum of $Q(n, m)$, with weights once again based on local image saliency. Based on $Q_W$, $\mathbf{Q_E}$ index was introduced as

$$Q_E = Q_W(X_1, X_2, F)^{1-\alpha} Q_W(X'_1, X'_2, F')^{\alpha} \quad (12)$$

where $\cdot'$ denotes gradient magnitude image. For $\alpha = 0$ it is simply equal to $Q_W$, but $Q_E$ allows to include gradient information and clearly becomes a purely gradient-based method for parameter $\alpha$ equal to 1, which was used in the experiments presented further.

### 3.4 Subjective Evaluation

Subjective evaluation is performed by a human tester. For these investigations, fused images were shown in random order to a group of 15 testers, asked to sort the set of results according to the subjectively perceived quality. Some of the results were shown more than once (in different random order) to address the question of reproducibility of the subjective evaluations.

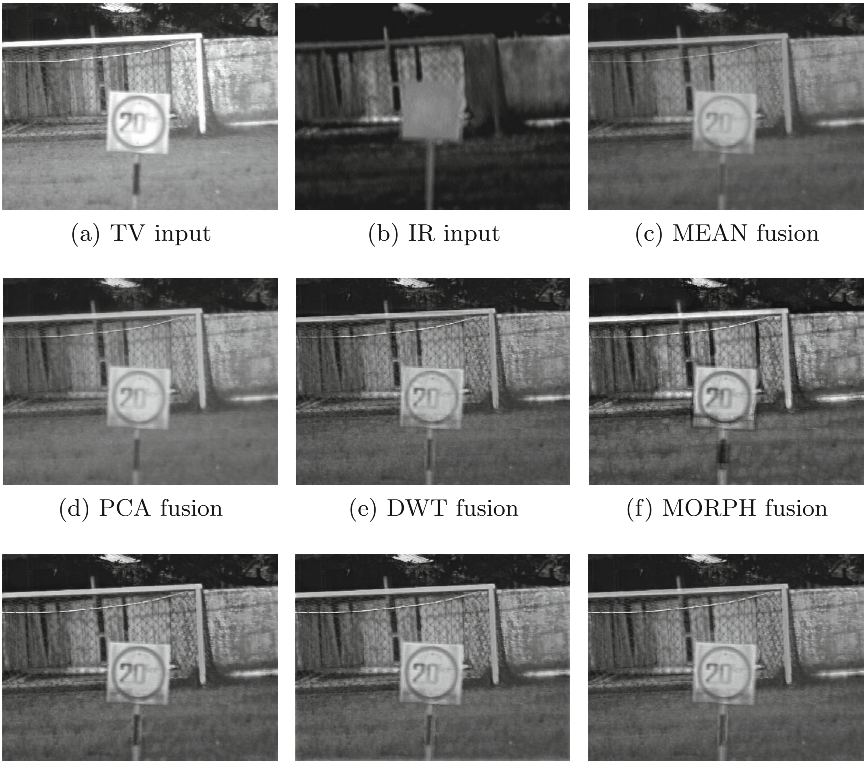
## 4 Results

In Tab. 1 results of evaluations performed with described measures are given, along with the time of calculation rankings for both fusion algorithms and quality measures. The best result for every measure is bold-faced. Corresponding input data and fused images are shown in Fig. 2. The chosen image is a representative part of the larger data set on which quality measures and fusion algorithms were tested.

Note that MEAN method is rated as the worst by both gradient-based methods, $Q_E$ and $Q_W$, while it is preferred by statistical measures.

**Table 1** Evaluation results

| | $H$ | $M_F$ | $M_F^\alpha$ | $M_F^{JS}$ | $Q^{AB/F}$ | $WGE$ | $Q_W$ | $Q_E$ | Subj. (rank.) | Speed (rank.) |
|---|---|---|---|---|---|---|---|---|---|---|
| MEAN | 7.337 | 3.655 | 32.66 | **0.539** | 0.477 | 4.361 | 0.743 | 0.654 | 7 | **1** |
| PCA | **7.441** | **3.899** | **32.92** | 0.538 | 0.557 | 3.341 | 0.824 | 0.811 | 5 | 2 |
| DWT | 7.223 | 2.591 | 6.926 | 0.497 | 0.589 | 2.915 | 0.786 | 0.840 | 4 | 5 |
| MORPH | 7.093 | 2.499 | 7.030 | 0.423 | 0.589 | 3.201 | 0.794 | 0.792 | 6 | 7 |
| LAPL | 7.309 | 2.627 | 7.427 | 0.513 | **0.634** | 2.533 | **0.831** | **0.882** | **1** | 3 |
| FAB | 7.231 | 2.545 | 6.564 | 0.457 | 0.614 | **2.498** | 0.817 | 0.851 | 2 | 4 |
| SIDWT | 7.222 | 2.691 | 7.975 | 0.500 | 0.621 | 2.609 | 0.817 | 0.863 | 3 | 6 |
| Speed (rank.) | 1 | 6 | 5 | 3 | 4 | 2 | 7 | 8 | - | - |



(a) TV input (b) IR input (c) MEAN fusion

(d) PCA fusion (e) DWT fusion (f) MORPH fusion

(g) LAPL fusion (h) FAB fusion (i) SIDWT fusion

**Fig. 2** Exemplary fusion results

## 5 Conclusions

Gradient-based algorithms and $Q_W$ measure correspond much better to the subjective evaluations than presented statistical methods. $WGE$ measure, proposed here, can be calculated relatively fast in comparison with other presented methods (cf. Tab. 1, last row), therefore it could possibly be well suited for the automatic real-time video fusion quality control application. FABEMD-based fusion algorithm was found to be one of the most effective, altogether with the Laplacian pyramid and SIDWT. Interestingly, a large correlation between subjective evaluation results was observed when the same results were shown to the testers twice (in different random order). Testers were often able to reproduce the same marks and not a single case of essentially contradictious evaluations was observed. This shows that the subjective evaluation is reproducible (for a given person, obviously) and therefore there is a certain meaningfulness to this kind of measure.

**Acknowledgement.** This work is supported in part by the research project No O R00 0019 07 of National Centre for Research and Development in Poland.

## References

1. Putz, B., Timofiejczuk, A., Bartys, M., Gwardecki, J.: System of TV and thermal image fusion for real-time application monitoring. Pomiary Automatyka Kontrola 57(7), 784–788 (2011)
2. Putz, B., Bartyś, M., Antoniewicz, A., Klimaszewski, J., Kondej, M., Wielgus, M.: Real-time Image Fusion Monitoring System: Problems and Solutions. In: Choraś, R.S. (ed.) Image Processing & Communications Challenges 4. AISC, vol. 184, pp. 147–156. Springer, Heidelberg (2012)
3. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Processing Letters 9(3), 81–84 (2002)
4. Piella, G.: New quality measures for image fusion. In: Proc. 7th Int. Conf. on Information Fusion, pp. 542–546 (2004)
5. Mitchell, H.B.: Image Fusion: Theories, Techniques and Applications. Springer (2010)
6. Rockinger, O.: Image sequence fusion using a shift invariant wavelet transform. In: Proc. IEEE Int. Conf. Image Processing, vol. 13, pp. 288–291 (1997)
7. Ahmed, M.U., Mandic, D.P.: Image fusion based on Fast and Adaptive Bidimensional Empirical Mode Decomposition. In: Proc. IEEE 13th Int. Conf. on Information Fusion, pp. 1–6 (2010)
8. Wielgus, M., Antoniewicz, A., Bartys, M., Putz, B.: Fast and Adaptive Bidimensional Empirical Mode Decomposition for the Real-time Video Fusion. Accepted at IEEE 15th Int. Conf. on Information Fusion, pp. 1–6 (2012)
9. Bhuiyan, S.M., Adhami, R.R., Khan, J.F.: A novel approach of fast and adaptive bidimensional empirical mode decomposition. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 1313–1316 (2008)
10. Zhang, X., Chen, Q., Men, T.: Comparison of fusion methods for the infrared and color visible images. In: Proc. 2nd IEEE Int. Conf. on Computer Science and Information Technology, pp. 421–424 (2009)
11. Qu, G., Zhang, D., Yan, P.: Information measure for performance of image fusion. Electronics Letters 38, 313–315 (2002)
12. Cvejic, N., Canagarajah, C.N., Bull, D.R.: Image fusion metric based on mutual information and Tsallis entropy. Electronics Letters 42, 626–627 (2006)
13. Zheng, Y., Qin, Z., Shao, L., Hou, X.: A Novel Objective Image Quality Metric for Image Fusion Based on Renyi Entropy. Information Technology Journal 7, 930–935 (2008)
14. Xydeas, C., Petrovic, V.: Objective image fusion performance measure. Electronics Letters 36, 308–309 (2000)
15. Jamrozik, W., Fidali, M.: Estimation of image fusion methods for purposes of vision monitoring of industrial process. In: Proc. of Diagnostics of Processes and Systems, pp. 459–464 (2011)

# Optimization of Memory Accesses for CUDA Architecture and Image Warping Algorithms

Przemysław Mazurek

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

**Summary.** The CUDA software platform gives abilities of outstanding performance for parallel computing using GPGPUs (General Purpose Graphics Processing Unit). The external memory interface is the main bottleneck of GPGPU for memory intense operations. There are a lot of reduction ways of this disadvantage for real–time applications. The profiling of the algorithm and execution parameters fitting are presented as a solution for the minimization of execution time. The fisheye to perspective transform is optimized as the example of the nonlinear image warping algorithm. The code optimization using search of the optimal kernel starting parameters is necessary. Such optimization gives better results for all cases due to limited processing area and the execution time is about 12% smaller. The unconventional method for CUDA of block–to–image assignment is emphasized.

## 1 Introduction

Image processing algorithms are computational power demanding for a lot of applications. Image compression and pattern recognition are well recognized areas, where high performance computational power is necessary. There are a numerous processing architectures important for parallel processing of images. The SIMD (Single Instruction, Multiple Data) processors, multiple cores architectures and GPGPU's are available.

Real–time video processing using modern and fast CPUs is possible. Reduction of the computation time would be obtained by proper design of the algorithm, also. Development of algorithms for specific software and hardware platforms is the optimal solution. Dedicated architecture has own processing limits usually, that should be considered for computation time reduction.

The optimization technique for GPGPU is shown in this paper, with attention to the memory access reduction. One of interesting algorithms is the image rectification from fisheye distorted image [4, 10] to rectilinear projection. Such algorithm is necessary in the real–time robot vision, video surveillance, and panorama stitching [3].

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 127–134.
springerlink.com 

## 2 Fisheye to Rectilinear Conversion Algorithms

Image warping algorithms are in a wide class of nonlinear image transforms, related to the spatial component of image [9, 10]. They are applied to the correction of image distortions, morphing, and special 2D effects [11]. There are two techniques used for the warping: non–parametric and parametric [6].

Non–parametric technique uses dense 2D vector fields for the description of desired pixels movements. This technique is memory consuming, due to the vector field storage, but gives abilities of description of complex warping operations. The vector field is also stored in memory, so additional fetching is necessary. The vector field have own resolution and large memory area is required (about image size).

Mathematic formulas are also applied for the description of typical distortions. The parametric technique is considered in this paper and is controlled by the limited set of parameters. The memory consumption is lowest, but the computation time is extended.

Both techniques are used for the correction of image distortions, introduced by fisheye lenses (Fig. 1).

**Fig. 1** Original distorted image (left) and after correction (right)

Fisheye lenses give a very wide angle of view, with significant distortions. Real lines are curves in the acquired image, for example. Fisheye lenses are very important due to very large of deep–of–field, so continuous focus control is not required.

Warping algorithm would be defined in two ways. First way is based on the calculation of the pixel's coordinate in the destination image using the source pixel coordinates. The second way is opposite, based on the calculation of the pixel's coordinate in the source image using the destination pixel coordinates. [2].

Both ways are valuable and they have own advantages and disadvantages. The first way is considered in this paper, but both of them would be derived from the following formulas [1]:

$$R_f = 2f \sin\left(\frac{\theta}{2}\right) \tag{1}$$

$$R_p = f \tan(\theta) \tag{2}$$

$$\frac{x_p}{y_p} = \frac{x_f}{y_f} \tag{3}$$

where:
$f$ – focal length,
$\theta$ – angle of movement vector,
$R_f$ – radial position of a pixel on the fisheye image from the center,
$R_p$ – radial position of a pixel on the perspective image from the center,
$(x_p, y_p)$ – destination pixel in the output image (corrected image),
$(x_f, y_f)$ – source pixel in the input image (distorted image).

Pixel location in the source image is obtained after eliminating $\theta$ from formulas. The radial position is the Euclidean distance between two locations, particular pixel coordinate and the center of the image:

$$R_* = \sqrt{(x_* - x_{*C})^2 + (y_* - y_{*C})^2} \tag{4}$$

Both methods are nonlinear spatial operations and influence overall performance due to the memory accesses.

## 3 Optimization of the Memory Accesses – Algorithm Level

This optimization technique is related to the algorithm [5] with attention to the implementation details, depended on the particular software/hardware platform, e.g. CUDA. Not all pixels in the destination image pixels should be processed, what is shown as the black areas in the corrected image Fig. 1. This straightforward technique would be applied by the introducing the mask (Fig. 2) to the algorithm memory accesses.

The direct implementation of the mask is not feasible for CUDA platform. Additional memory area is necessary due to the limited bandwidth of the

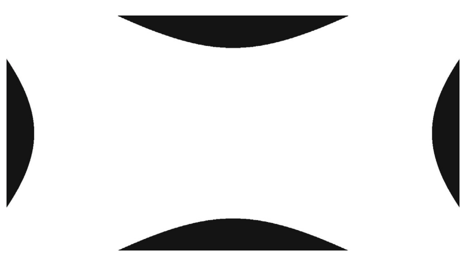


**Fig. 2** Output image mask used for active area selection (white)

memory interface. Another approach is applied – the Look–Up Table (LUT) for the storage of top–left corner coordinates of the CUDA processing blocks. Every corner position is stored as two values $X$ and $Y$ in LUT.

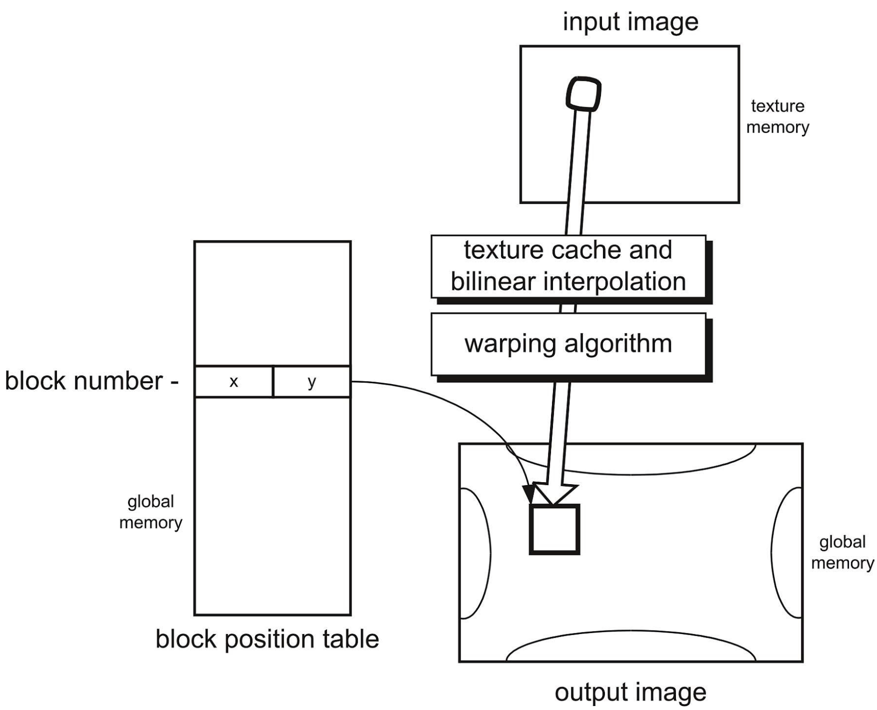



**Fig. 3** Schematic of processing

The processing block corresponds to the appropriate destination image area and blocks should not overlap (7). The LUT is filled before image processing and simple optimization technique is applied for block to the mask $M$ assignments [5]:

$$E = \min_i (B_i) \tag{5}$$

$$\sum_i B_i \supset M \tag{6}$$

$$B_j \not\supset B_i, i \neq j \tag{7}$$

where E is the error value that should be minimized to the zero. The filling of LUT is not the subject of the GPGPU. The LUT is fixed for particular lens and is filled by the e.g. Matlab and incorporated into GPGPU kernel code.

The optimization algorithm starts from the left column of the mask. All columns have fixed width. Every row in the column is marked, if there is one or more of the input image related pixels in this row. The vertical position of the first one available marked row, that is not assigned yet, is used as a Y–value of the LUT. The X–value is calculated from the particular column number. The next rows (marked or not) are also assigned to this block and the number of them depends on the assumed height of the processing block.

An example assignment of blocks are shown in Fig. 4 for the description of the algorithm. The first block 'A' is assigned to top–left corner. Some part of the block are not important for processing (top–right), but other parts are necessary. The position of the block is stored in LUT. The next block 'B' is located under previous block. There are some parts of image that should be processed so position of this block is stored in LUT, and this process is repeated in vertical direction. The next column of blocks should not start from first row, because this area is not important. The first pixel that should be processed is located in the top–left part of the block 'C'. Some part of block 'C' is not important for computation. Again, a next block in vertical position is considered 'D'. In the third column situation is repeated, and so on.

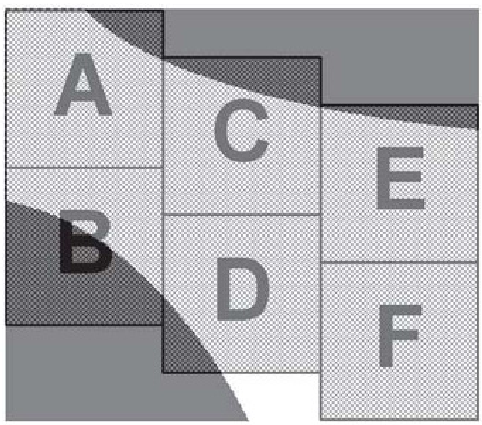



**Fig. 4** Example of top–left corner with block markings

The example assignment is shown in Fig. 5 and corresponding LUT defines the processing area.

There are two variables: the width and the height of block and optimal values would be determined by the profiling only.

In this example the input image has 640x428 size and the output image has 1280x664 size. The output image is magnified additionally, what is useful option, because the output image has a nonlinear optical resolution.

**Fig. 5** Processed (white) and omitted (black) areas (height=16, width=1)

## 4 GPGPU Architecture

GPGPUs have great impact on parallel processing, and they are derived from shader language processing units (GPUs). They are available in a wide range of computation power devices. The parallel code for Nvidia CUDA engine [7, 8] is developed and processed using SIMT (Single Instruction Multiple Thread) model using C–language. Threads, that execute the kernel code are organized as a grid, filled by the thread blocks. Multiple blocks are processed in parallel, depending on the free resources of GPGPU. The global memory is used for storage of output image and the texture memory is used for storage of input image, in case of the image warping. Both of them are located in the external GPGPU memory, what is significant disadvantage due to the bottleneck (bandwidth and latency) of global memory. The texture memory is accessed by the internal on–chip cache with the additional 2D interpolation unit what is useful for the image processing algorithm like image warping.

## 5 Optimization of the Memory Accesses – Runtime Profiling

Different block sizes should be used depending on the hardware properties of particular GPGPU. The example results are related to the Nvidia's G80 (GeForce 8800GTS) and are different for another cards. Using early profiling – testing of the execution time for different configurations of blocks, the optimal values are obtained and used later in real–time operation. A mean value should be calculated for every block size.

There are recommended block widths – multiples of the number 16, because every pixel occupies single byte, and this is the optimal coalescence value for this GPGPU [7, 8]. As the illustration of the non–optimal widths the following values are tested: 8, 12, and 24 additionally. The results of LUT–based image processing without applied algorithm's optimization are shown in Fig. 6. The optimization (Fig. 6) gives better results for all cases due to limited processing area - the execution time is about 12% smaller. The single

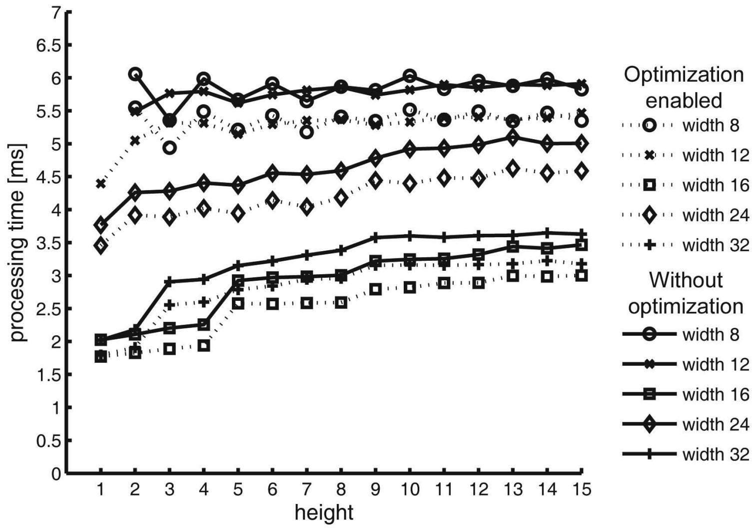



**Fig. 6** Performance of algorithm – without and with proposed optimization

row (*height* = 1) is optimal for the best widths (*width* = 16 or 32). There are significant differences (about 40% of the value) between small and medium height sizes, visible as steps and it could be interpreted as a performance reduction of texture cache due to limited size and reduced hit–to–miss ratio.

## 6 Conclusions

The profiling is necessary due to destination GPGPU, because optimal block size can not be determined using algorithm analysis only. The algorithm analysis with the special care about coalescence is possible for more regular memory accesses, where the input and output image blocks correspond directly (fixed and equal block sizes, equal address offset of the image blocks). Such correspondence is not available for the image warping operations, so optimal code with utilization of the shared memory is very hard to design and depends on the correction parameters. The application of the texture cache and two levels of optimizations gives the ability of the algorithm development at very short time.

Presented method of the memory accesses is different in comparison to the typical implementation of CUDA code in the image processing applications. CUDA manuals and books prefer 2D block grid access to the image operations, and presented method is based on 1D block grid with LUT. The LUT values are converted to the XY image coordinates.

**Acknowledgement.** This work is supported by the MNiSW grant N514 004 32/0434 (Poland). 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. Altera: A Flexible Architecture for Fisheye Correction in Automotive Rear–View Cameras, Altera White Paper (2008)
2. Bradski, G., Kaehler, A.: Learning OpenCV. In: Computer Vision with the OpenCV Library. O'Reilly (2008)
3. Dersch, H.: Panorama Tools. In: International VR Photography Conference on Open Source Software for Immersive Imaging, Berkeley (2007), http://webuser.fh-furtwangen.de/~dersch/IVRPA.pdf
4. Alvarez, L., Gomez, L., Sendra, J.R.: Algebraic Lens Distortion Model Estimation, Image Processing On Line (2010), http://dx.doi.org/10.5201/ipol.2010.ags-alde
5. Mazurek, P.: Real–Time Correction of Cameras' Geometric Distortions using GPGPU. In: 14th IEEE/IFAC International Conference on Methods and Models in Automation and Robotics, MMAR 2009, Miedzyzdroje (2009)
6. Mazurek, P.: Mobile system for estimation of the internal parameters of distributed cameras. Measurements Automation and Control 56(11), 1356–1358 (2010)
7. NVIDIA: CUDA – Compute Unified Device Architecture. C Programming Guide (2012), http://developer.download.nvidia.com/compute/DevZone/docs/html/C/doc/CUDA_C_Programming_Guide.pdf
8. NVIDIA: CUDA - Compute Unified Device Architecture. Reference Manual v4.0, Nvidia (2012), http://developer.download.nvidia.com/compute/DevZone/docs/html/C/doc/CUDA_Toolkit_Reference_Manual.pdf
9. Russ, J.C.: The Image Processing Handbook. CRC (2006)
10. Solomon, D.: Transformations and Projections in Computer Graphics. Springer (2006)
11. Velho, L., Frery, A.C., Gomes, J.: Image Processing for Computer Graphics and Vision. Springer (2009)

# Emulation of the Double-X Cinematography Negative Film

Rafał Stegierski and Karol Kuczyński

Maria Curie Sklodowska University, Institute of Computer Science
rafal.stegierski@gmail.com

**Summary.** Article presents work-flow which gives possibility to emulate characteristics of the one of the most popular black and white film negative used in cinematography. Presented procedure is based on examination of the source material and technical data available from manufacturer and could be easily used for preparation characteristics of other available stocks.

## 1 Premises

Emulation of the analogue photography materials in digital photography is doubtful because still wide variety of materials is available, prices of it still are on the acceptable level and equipment is easy to obtain. In case of the cinematography, especially independent, this procedure is more justified. From negative material through laboratory to camera, everything is very expensive at each stage of the production. In some cases, such as historical documentary productions, there are also need to match archival footage taken with such a material and modern, video recordings. Obtaining analogue film look in digital domain is much more complicated than simple desaturation of the image because it should take into account a lot of different factors due to nature of the both analogue and digital material. Main purpose of emulation of the analog film stock is possiblity to obtain similar feel and look in situation when additional founds for materials and process is unavailable. In such a situation erzac is much better than nothing.

## 2 Film Stock Preparation

Eastman Double-X is high-speed, general production negative black and white cinematographic film still available in 16 mm and 35 formats. It's exposure index is equal to 250 ISO in daylight [1].

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 135–142.
springerlink.com 

Eight frames of the film was exposed with step equal to 1EV with defocused lens and lightened with cold cathode diffused lamp. Then material used for characteristics preparation was developed in Kodak D-96 stock dilution. Film was developed 7 minutes in small tank in temperature of 21 Celsius degree then fixed with standard acid fixer in place of Kodak F-5 Fixer [2].

All frames of the developed film was scanned with dedicated 2K scanner in linear, and equal for each, mode without any post-processing including unsharp masking. For each of the frames average lightness level was calculated and characteristic curve was determinated. Its shape (fig. 1) is very similar to the rough one presented in film manufacturer documentation [1].

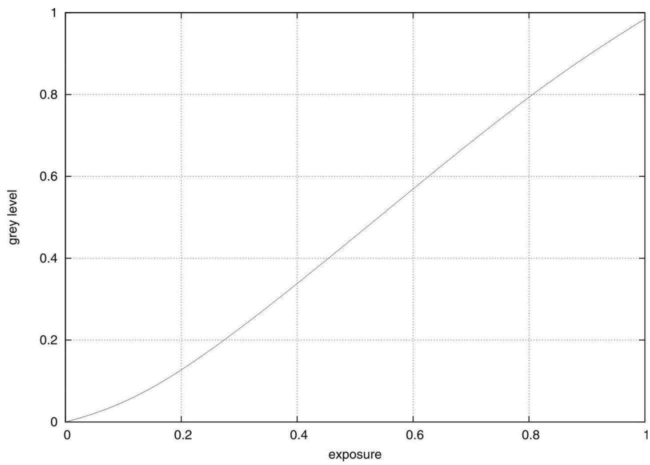



**Fig. 1** Contrast curve measured for Double-X negative

## 3 Video Material

In situation where video footage is manipulated in wide spectrum of changes should be taken in raw straight from image sensor without any contrast curve alteration. Only few, and rather expensive, cameras give us such a possibility. In most cases there is not only gamma correction but also knee setting at different IRE value usually from 80 to 100 units and different slope. In such a case calibration of the input image with, for example, standard Kodak/Tiffen Q-13 gray scale chart with twenty-zone grayscale values is needed.

Optimally video should be captured with 2K resolution for 35 or Super 35 emulation, but, because it is not standard resolution for video cameras, except dedicated to digital cinematography ones, any other format is acceptable. In this case resizing of the image should be make to match size of the frame of different formats and 2K scanner resolution. In most of the cases pixel aspect ration should be 1:1 if video was captured without anamorphic optics. This operation also lowering information in image which correspond with characteristics of the analogue material which has spatial frequency up to 100 line pairs per millimetre but in most cases it is under 50 lines what correspond to resolution of the scanner. After whole emulation process resolution could be changed to chosen production or emission standard. Proposition of the resolution for main film formats is in the table 1. Different, and today more often used, aspects ratio such as 1:1.87 or 1:2.39, are obtained just by cropping image in height.

**Table 1** Proposed resolution of the frames for different analogue cinematography formats

| Format | Height in pixels | Width in pixels | Aspect ratio |
|---|---|---|---|
| 8 | 308 | 418 | 1:35 |
| Super 8 | 372 | 538 | 1:1.45 |
| 16 | 696 | 954 | 1:1.35 |
| Super 16 | 690 | 1164 | 1:1.7 |
| 35 | 2048 | 1488 | 1:1.37 |
| Super 35 | 2320 | 1690 | 1:1.37 |

Next one thing is recommendation to capture image with native speed of sensor without any gain because it could place unwanted amount of digital noise in data, increase overall contrast of capture and fall shadows.

## 4 Colour and Contrast Transfer

Eastman Double-X has non linear spectral response. The highest sensitivity is between 400 and 450 nm of wavelength than it decreases and between 500 and 630 nm is rather constant. Over 630 nm spectral sensitivity fall down rapidly. Main problem with converting YUV or RGB information stored in video data captured with camera is fact that there is no straight information about wavelength and resultant color derives from distribution of wavelength and light power. To mimic spectral response of the analogue material we propose to find dominant wavelength in captured color. Presented method give similar results to calculation of the Hue component in HSL or HSV colour models but works in RGB space and take account into different distribution of the components.

All components in each used color space in this article are normalized and treat as real numbers. Wavelength is expressed in nanometres.

All possible combinations of the colour in RGB space are placed in cube. Achromatic colours are on the diagonal from point (0, 0, 0) to (1, 1, 1).

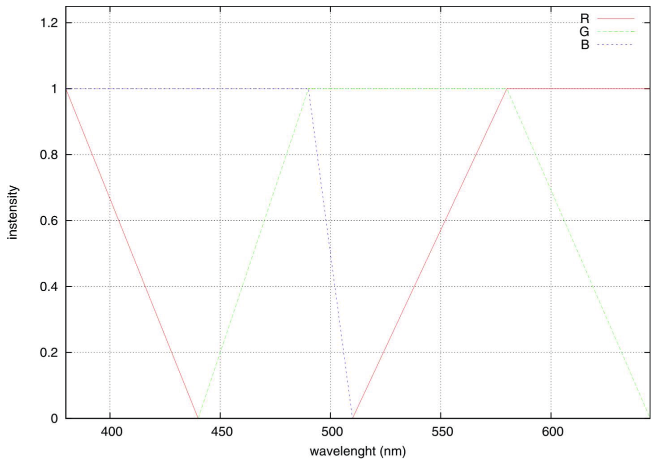



**Fig. 2** Contrast curve measured for Double-X negative

Combinations of the RGB components which gives monochromatic colours in correspondence to wavelength from 380 nm to 645 nm (see figure 2) are located on the edges the cube. Calculation of dependence between wavelength and colour is based on Dan Burtons [3] code. Dominant wavelength is find out with function $f_w(r, g, b)$. It uses mentioned earlier wavelength to RGB $f_{rgb}(w)$ mapping function. Values $r$, $g$, $b$ components of the point positioned on the line of monochromatic colour when distance in Euclidean distance from pixel colour $p$ in RGB space is minimal.

$$\lambda = f_w(r, g, b) \quad for \quad \{r, g, b : min||f_{rgb}(w) - p||\} \tag{1}$$

Because spectral response affects lightness of the pixel there is defined function $L_w(\lambda)$ which represents lightness multiplier for pixel according to wavelength. In software implementation this function is released as a look up table with course as presented on figure 3.

Lightness and saturation are calculated as has it place in HSL model [4].

To avoid problems with smooth transitions for colours defined near achromatic final lightness for each pixel is calculated with correspondence to saturation $S$ as (2) where $L_a$ is average lightness correction. If $L_w(\lambda)$ is implemented as a LUT it could be calculated as a arithmetic average of elements in the table.

$$L_a = \frac{\int_{380}^{645} L_n(\lambda) d\lambda}{265} \tag{2}$$

$$L_n(\lambda) = L_0 * (L_w(\lambda) * S + L_a * (1 - S)) \tag{3}$$

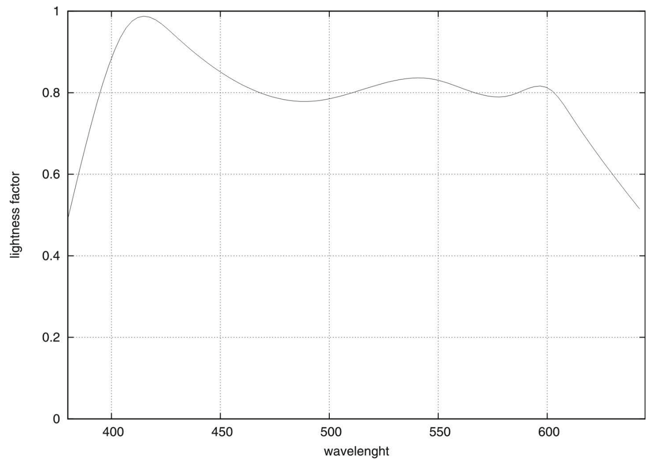



**Fig. 3** Shape of the $L_w(\lambda)$ function

In next step contrast correction is made with function $L_c(L)$ derived from measured contrast curve (figure 1). If video footage used as a input has non linear contrast characteristic function has to be a composition of inverse function for contrast characteristic of input image and correction. Lightness for each pixel is represented as (4).

$$L_{nc}(\lambda) = L_c(L_n(\lambda)) \tag{4}$$

## 5 Grain

One of the most prominent differences between analogue and digital captured image is noise characteristic. In matter of fact we should rather use term grain in case of analogue. Its cause is a fact that image is build with different in size and distribution particles of metallic silver. To obtain information about grain in case of Double-X for each of the exposures histogram of the image was made. In result information about near Gaussian distribution and independence of grain and exposure is find out. According to this information unified grain map was created in process of shifting values of pixels in grain image with formula (5) where $G_a$ is average, $G_m$ lightness of pixel for whole image.

$$G_n = G_0 - G_a \tag{5}$$

To avoid semi static temporal characteristic of the grain for each on of the frames position and rotation of the map is generated pseudorandomly.

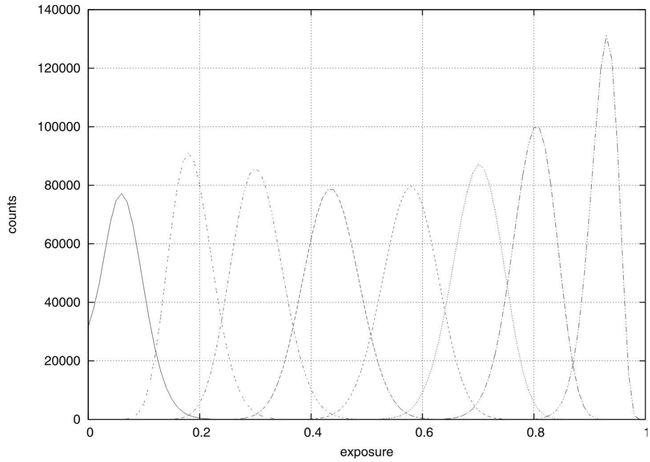



**Fig. 4** Histogram of grain characteristics for eight exposure values

Final lightness of the pixels is calculated as (6). Because emulated film stock is black and white saturation for whole image is zeroed.

$$L_{ncg} = G_n + L_{nc}(\lambda) \tag{6}$$

## 6 Results

Figure 5 presents one frame of the input capture, simply desaturated version, and version generated with method presented in the article. Images has different tonality and characteristic. There is also changes in representation of patches at Gretag Macbth Target. Figure 6 is input image and converted version of the outdoor scene.

In near future more tests are planed with direct comparison emulated image and Double-X footage.

## 7 Conclusions

Presented in this article method is rather suggestion of the possible workflows than final recipe. Emulation is build around real analogue material and its analysis in domain of spectral response, contrast and grain. To bypass problem of lack a way to find out spectral characteristic of captured light method of designation of dominant wavelength and proper representation of spectral response for near achromatic colors was proposed. Further works and more test are necessary.

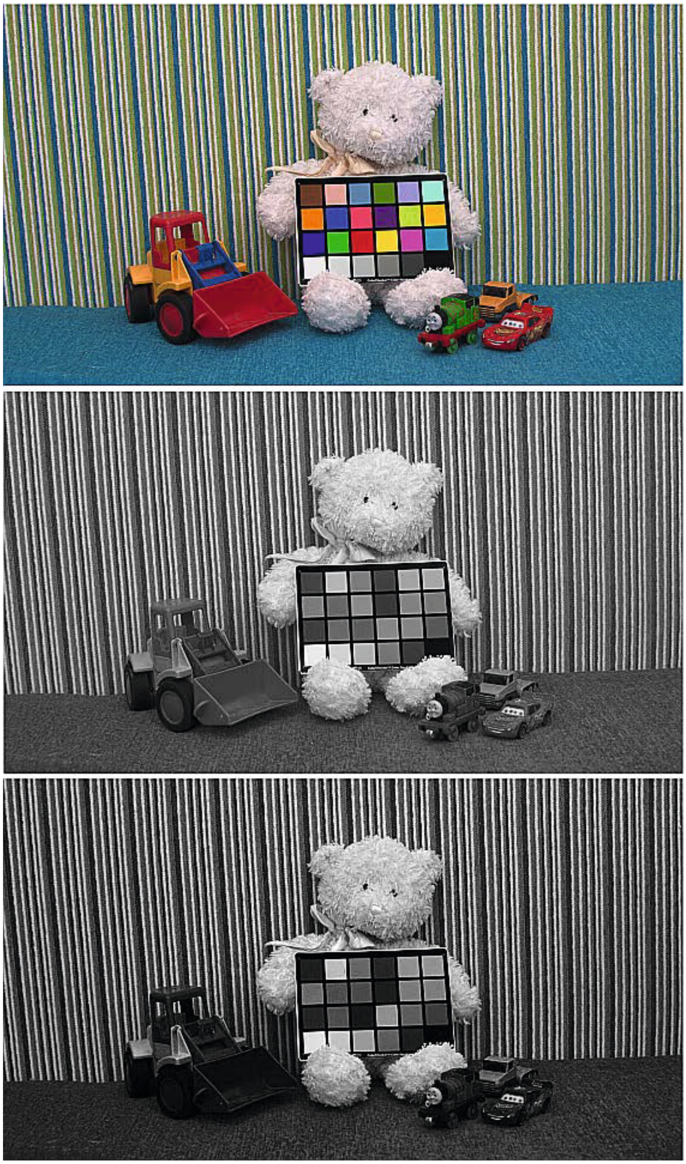


**Fig. 5** Input image, desaturated, Duble-X emulation

**Fig. 6** Input image, Duble-X emulation, outdoor scene

## References


1. Eastman Double-X Negative Film 5222$^{TM}$/7222$^{TM}$ Technical Data. Kodak (1998)
2. Processing Kodak Motion Picture Films, Module 15 Processing Black-and-White Films, Technical Data. Kodak (2000)
3. Burton, D., http://www.midnightkite.com/color.html (1996)
4. Joblove, G.H., Greenberg, D.: Color Spaces for Computer Graphics. In: Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques (1978)

# Real-Time Image Fusion Monitoring System: Problems and Solutions

Barbara Putz[1], Michał Bartyś[1], Adrian Antoniewicz[1], Jan Klimaszewski[1], Marcin Kondej[1], and Maciej Wielgus[2]

[1] Institute of Automatic Control and Robotics,
ul. św. Andrzeja Boboli 8, 02-525 Warsaw, Poland
{bputz,bartys}@mchtr.pw.edu.pl, aantoniew@gmail.com,
janklimaszewski@wp.pl, marcin@kondej.eu

[2] Institute of Micromechanics and Photonics,
ul. św. Andrzeja Boboli 8, 02-525 Warsaw, Poland
maciek.wielgus@gmail.com

**Summary.** The paper presents complete real-time image fusion system intended for supervisory and monitoring purposes in mobile appliances. The system is equipped with an optoelectronic head containing two multiband image sensors: TV and IR (infrared). System structure as well as searching process of the efficient, real-time image processing algorithms are briefly described. Principally, image registration is based on the hybrid approach (edge extraction and phase correlation). The image fusion method can be chosen according to operator demands between Laplacian pyramid and FABEMD decomposition. Real-time multispectral video signals processing is performed by application of custom hardware solution based on a single FPGA chip. This allows for implementing fast paralleled and pipelined processing flow. The system hardware is also presented in this paper.

## 1 Introduction

Contemporary integrated electronic heads may consist of several cameras sensitive to multiple wavebands. Typical example is an assembly (head) of TV camera combined with IR one (Fig. 1). Video system making use of such a head is usually equipped with a single display. Thus the operator needs to switch over both cameras dependently of atmospheric and other conditions. The image fusion system of multimodal images is proposed in order to avoid necessity of switching cameras. Such a system is currently being realized with the support of a research project No O R00 0019 07 of National Centre for Research and Development in Poland by the consortium of Warsaw University of Technology, Silesian University of Technology and Bumar Elektronika S.A. The prototype of the system, working in real-time and named UFO (Uklad Fuzji Obrazów - in Polish) is the aim of the project [1], and is foreseen for industrial implementations. The paper presents the problems and solutions

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 143–152.
springerlink.com 

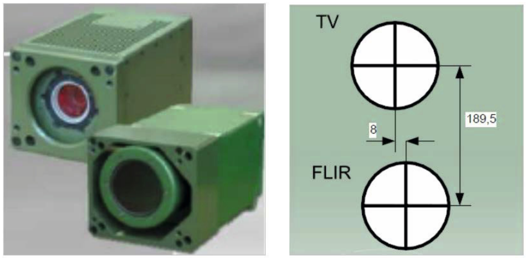



**Fig. 1** TV and IR cameras and their arrangement in optoelectronic head

of image registration (Section 2) and image fusion (Section 3). As the result, the system structure (Section 4) and the principles of FPGA implementation of the system (Section 5) are presented. Finally, hardware solution of UFO (Section 6) as well as conclusions are presented.

## 2 Image Registration

The component image alignment, often known as image registration, is the most challenging task in overall fusion process, significantly difficult in case of multimodal image pairs. The image misalignment arises from different optical sensor paths (Fig. 1), different sensors' resolution, field of view, camera distortions etc. General geometric considerations lead to necessity of using a projective image warp. To obtain optimal image alignment it is necessary to calculate the set of projective transform coefficients. The typical accurate multimodal registration (as performed for example for medical images) cannot be realized in real time, because it involves optimization techniques of mutual information or other measure; moreover, they can lead to local, not global extreme [2, 3].

The good performance of image registration can be achieved for single modal images thanks to fast phase correlation methods working with the use of Fast Fourier Transform. With accompanying log-polar transform, the coefficients of rigid body alignment (translation, rotation, scaling) can be obtained and this type of registration is entirely sufficient in most practical real-time applications.

During realization of UFO project, the phase correlation method has been tested and considered as image registration method used in non-standard way, i.e., for multimodal images. The subsequent research proved that it is possible to obtain reasonable effects for multisensor registration with the use of image edges extraction as pre-processing step before starting phase correlation algorithm. The several edge detectors (Harris, Canny, morphological) have been simplified or improved and tested with surprisingly good results of translation coefficients. The details are presented in a separate paper [4].

Although the final release of UFO video system will be used on mobile platform with IR and TV camera head, it has been proved that change of rotation coefficients during the work can be entirely neglected due to the assumption of cameras arrangement as in Fig. 1 and their rigid, stable position in the head. During tests of the head, the small offset of rotation angle (below 0.7 degree) between both cameras was observed, but certainly it has no noticeable influence on results of image fusion. In addition, significant variance value of calculated angles for image sequences could be very annoying for system operator. Finally, we decided to consider - as additional to translation coefficients - only scaling factors between aligned images, which are calculated one-off, as constant values, on the base of cameras parameters.

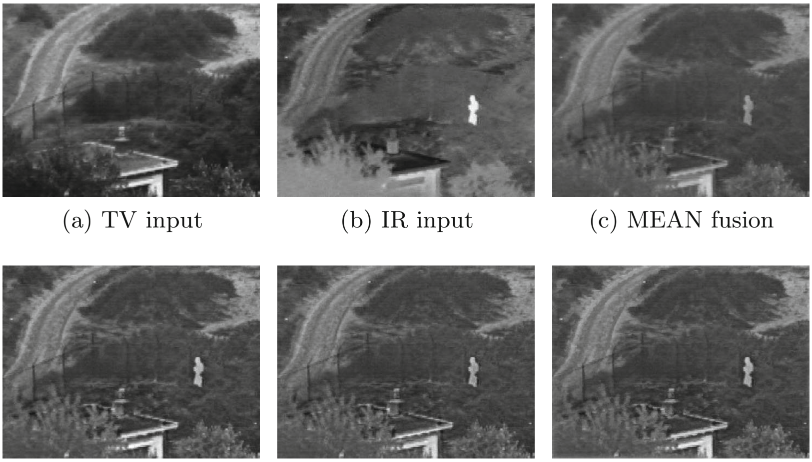
## 3 Image Fusion

The goal of TV and IR image fusion is to combine the complementary information from each camera into a single, enhanced, high quality image with details extracted from both component images. The simple fusion by averaging (Fig. 2c) cannot generally produce satisfactory results, particularly in situations, where some details are represented only in one image (e.g. the man's figure in Fig. 2b or the chimney in Fig. 2a) [1]. Fusion by averaging destroys contrast in such cases and the typical solution is the use of multiresolution, decomposition methods [5] which enhance and extract image features without loss of contrast (Fig. 2d-2f).

The fusion algorithm for real-time image fusion application has to meet several basic requirements: effective data processing in time limited to few milliseconds, should allow for smooth fused video output, output satisfactory and predictable results allowing for unsupervised system work in most cases, and providing algorithm complexity allowing for direct low-level implementation. From a larger group of fusion methods, three were chosen based on their performance. These are: Laplacian pyramid method (Fig. 2d), shift invariant wavelet transform method (SIDWT, Fig. 2e) and an algorithm based on the fast and adaptive bidimensional mode decomposition (FABEMD, Fig. 2f). Additionally, simple fusion by averaging was selected for comparison purposes (Fig. 2c). Except for the FABEMD fusion algorithm, recently introduced in [6] and adopted for the real-time image fusion in [7], these are all well established methods. FABEMD is an accelerated variant of the bi-dimensional empirical mode decomposition algorithm, for which high quality fusion results were demonstrated prior, e.g. [9] with experimental results for the same images, as in Fig. 2.

Exemplary fusion results shown in Fig. 2 were evaluated with several measures of quality suited for image fusion purposes (Tab. 1). The measures used

[1] Figs. 2a and 2b from TNO, The Netherlands [8], http://www.imagefusion.org



(a) TV input (b) IR input (c) MEAN fusion

(d) Laplacian fusion (e) SIDWT fusion (f) FABEMD fusion

**Fig. 2** Exemplary fusion results

are mutual information MI [10], objective image fusion performance measure $Q^{AB/F}$ [11] and weighted gradient error WGE. Except for MI, which was reported to favor simple averaging, evaluations show similar performance of the selected methods, which is also clear from Fig. 2. Broader discussions and comments on real-time fusion quality evaluation are given in a separate paper [12].

**Table 1** Evaluation measures' results for different fusion methods

| | $MI$ | $Q^{AB/F}$ | $WGE$ | Subjective (ranking) | Speed (ranking) |
|---|---|---|---|---|---|
| Mean | 2.031 | 0.672 | 3.532 | 4 | 1 |
| Laplacian | 1.435 | 0.860 | 2.780 | 1 | 2 |
| SIDWT | 1.452 | 0.852 | 2.839 | 3 | 4 |
| FABEMD | 1.453 | 0.822 | 2.770 | 2 | 3 |

Based on quality evaluation performed on larger data set, time performance and algorithm potential for a fast FPGA implementation, Laplacian pyramid and FABEMD methods were selected for the UFO system.

### 3.1 The System Structure

As the result of presented deliberations and solutions, the simple structure of the UFO system has been established. Here, scaling and rotation parameters are fixed for IR image. The constant values of rotation, although not required in our system, can be designed, similarly as scaling factors, and taken on board. Image translations parameters are calculated in background of the real-time fusion process and are updated each consequent 3 image frames. Fig. 3 presents block schematics of the UFO system processing flow.

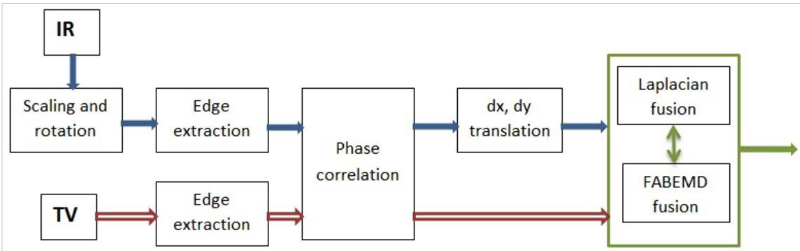



**Fig. 3** The block schematics of the UFO system

## 4 FPGA Implementation

The real-time image fusion monitoring system except image display and acquisition engine requires implementations of fusion and image alignment algorithms. To meet critical timing requirements for real-time system, highly paralleled and pipelined computation scheme should be used. In addition, fast and massive memory is required to handle buffer video frames and manage all of the system components. To meet all these requirements, an FPGA technology was applied. Primary, real-time video processing system was realized by means of Altera Cyclone III Video Kit (Fig. 4).

The structure of designed system is presented in Fig. 5. Components of FPGA capture images from video decoder into DDR memory. Frame reader sends output image from image buffer to output encoder. Access to memory is available by means of Avalon Bus which also provides arbitration process for system components. Image alignment algorithm is accelerated by implemented FFT/iFFT and Phase Correlations coprocessors linked with internal FPGA system bus. Harris edge detector was implement in order to achieve more robust image translation calculations. The main fusion process is performed by implemented FABEMD and Laplacian fusion components which are software selectable. Video system includes also NIOS software processor IP which is used for debugging purposes, non time-critical computations and for initialization of video codecs, etc. The structure of the on-chip system (Fig. 5) is an effective combination of FPGA performance and processor versatility. The FPGA resource requirements are presented in Tab. 2. Processing

**Fig. 4** The technology demonstrator of the UFO system

times of implemented algorithms are presented in Tab. 3. Complete video processing system allows for real-time image registering and fusion. The details of FPGA implementation of both fusion methods (FABEMD and Laplacian pyramid) are presented in a separate paper [13].

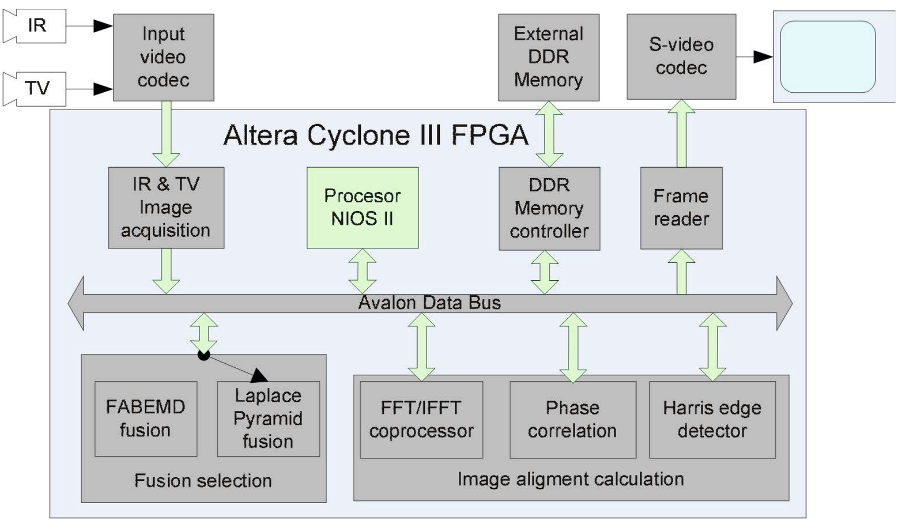



**Fig. 5** The structure of single FPGA-based image fusion system

**Table 2** FPGA resource requirements for fusion modules

| Resource type | Resource used | Total available |
|---|---|---|
| Logic cells | 26516 | 119088 |
| Registers | 11624 | 119088 |
| Memory bits | 771840 | 3981312 |
| DSP elements | 16 | 576 |

**Table 3** FPGA resource requirements for 640x480 images with 150MHz FPGA clock

| Component | Resource used | Total available |
|---|---|---|
| Image acqusition | Real-time | Preset to 40 ms |
| Fusion based on Laplacian Pyramid | Real-time | Preset to 40 ms |
| Fusion based on FABEMD | Real-time | Preset to 40 ms |
| Phase correlation image aligment method | 100 ms | * |
| Frame reader | Real-time | Preset to 40 ms |

## 5 Hardware Solutions

Multimodal real-time image registering and fusion system presented in this paper was primary intended for mobile and fixed platforms applications. In these cases, available off-the-shelf video development kits might be used exclusively for software development. Hence, we build-up an ăFPGA based custom system called UFO board that is sufficiently robust against harsh environments. Physically, UFO board has a form of 6U VME standard [14, 15] module. Layout of the developed module is shown in Fig. 6. As it was experimentally proven, UFO is able to manage real-time multimodal image fusion operating at 50 Hz and 60 Hz. The detailed structure of UFO system is presented in Fig. 7.

Analogue CCIR coded video signals from visible and infrared bands cameras are fed to multi-channel integrated NTSC/PAL video decoder via appropriate low band pass filters. Filters are build-up by means of a few passive discrete elements. All parameters needed for video signal processing such as hue, contrast, brightness, saturation and sharpness are programmable and are adjusted by means of I2C serial link with FPGA matrix. Video decoder generates digital video outputs and provides synchronization, blanking, lock and clock signals for FPGA. Luminance and chrominance of video outputs are coded according to ITU-R BT.656 specification. Digital video outputs are presented to FPGA. UFO board delivers monochrome images. Therefore, the only luminance data output is further used for processing. FPGA chosen for UFO prototype board is a low cost, low power device providing sufficiently large amount of logic elements, embedded memory, hardware multipliers, and user I/O's to meet multimodal image registration and fusion requirements. FPGA is running at 150 MHz clock frequency. Two banks of external DDRAM2 memories provide sufficiently space for begin, intermediate and final image buffers. Additionally, fast synchronous burst static RAM complete memory resources of the system. More detailed hardware considerations concerned with UFO system are presented in [16].

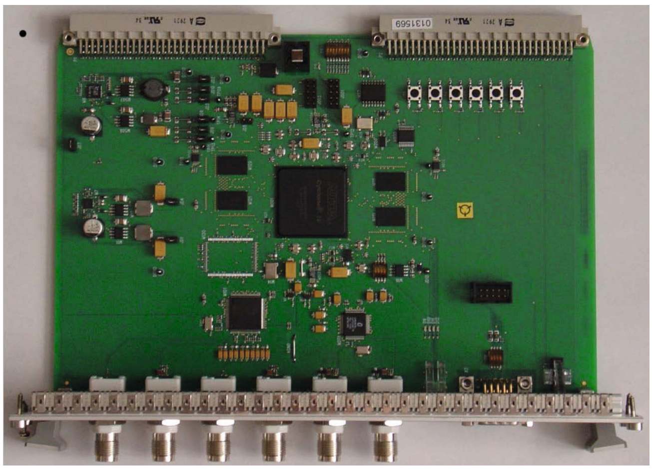


**Fig. 6** Layout of the UFO VME prototype board

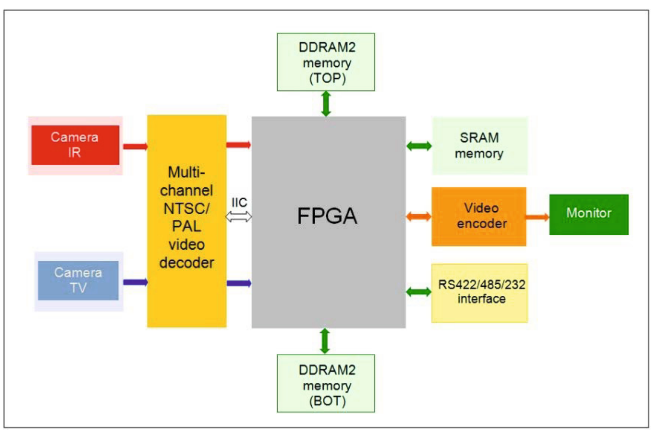



**Fig. 7** Simplified block schematics of the UFO prototype board

## 6 Final Remarks

Complete real-time multimodal image video processing system was presented in this paper. This system consists of hardware solution based on single FPGA chip and a software based of application of self developed IP components. Hardware solution, due its robust design allows for application in harsh environments as well as in mobile platforms. Duly designed and tested software makes possible to achieve real-time fusion and at least 10 Hz update rate of image registration.

It is worth to mention, that most of known and powerful image registering and fusion approaches are contemporary useless for real-time application due to their unacceptable computational complexity. Therefore, when developing UFO system, much effort has been spent to find out extremely fast and applicable image registering and fusion algorithms.

## References


1. Putz, B., Timofiejczuk, A., Bartys, M., Gwardecki, J.: System of TV and thermal image fusion for real-time application monitoring. Pomiary Automatyka Kontrola 57(7), 784–788 (2011)
2. Zitová, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing 21(11), 977–1000 (2003)
3. Wyawahare, M.V., Pradeep, M.P., Abhyankar, H.K.: Image Registration Techniques: An overview. International Journal of Signal Processing, Image Processing and Pattern Recognition 2(3), 11–27 (2009)
4. Klimaszewski, J., Kondej, M., Kawecki, M., Putz, B.: Registration of Infrared and Visible Images Based on Edge Extraction and Phase Correlation Approaches. In: Choraś, R.S. (ed.) Image Processing & Communications Challenges 4. AISC, vol. 184, pp. 157–166. Springer, Heidelberg (2012)
5. Dwyer, D., Smith, M., Dale, J., Heather, J.: Real time implementation of image alignment and fusion. In: Driggers, R.G., Huckridge, D.A. (eds.) Proc. SPIE Electro-Optical and Infrared Systems: Technology and Applications, vol. 5612, pp. 85–93 (2004)
6. Ahmed, M.U., Mandic, D.P.: fusion based on Fast and Adaptive Bidimensional Empirical Mode Decomposition. In: Proc. IEEE 13th Int. Conf. on Information Fusion, pp. 1–6 (2010)
7. Wielgus, M., Antoniewicz, A., Bartys, M., Putz, B.: Fast and Adaptive Bidimensional Empirical Mode Decomposition for the Real-time Video Fusion. Accepted at IEEE 15th Int. Conf. on Information Fusion, pp. 1–6 (2012)
8. Lewis, J.J., et al.: The Eden Project multi-sensor data set. Technical report TR-UoB-WS-Eden-Project-Data-Set. University of Bristol and Waterfall Solutions Ltd., UK (2006)
9. Liang, W., Liu, Z.: Region-based Fusion of Infrared and Visible Images Using Bidimensional Empirical Mode Decomposition. In: Int. Conf. on Educ. and Inf. Techn., ICEIT, pp. V3-358–V3-363 (2010)
10. Qu, G., Zhang, D., Yan, P.: Information measure for performance of image fusion. Electronics Letters 38, 313–315 (2002)

11. Xydeas, C., Petrovic, V.: Objective image fusion performance measure. Electronics Letters 36, 308–309 (2000)
12. Wielgus, M., Putz, B.: Comparative Analysis of Image Fusion Performance Evaluation Methods for the Real-time Environment Monitoring System. In: Choraś, R.S. (ed.) Image Processing & Communications Challenges 4. AISC, vol. 184, pp. 123–130. Springer, Heidelberg (2012)
13. Antoniewicz, A.: FPGA Implementation of Decomposition Methods for Real-time Image Fusion. In: Choraś, R.S. (ed.) Image Processing & Communications Challenges 4. AISC, vol. 184, pp. 167–174. Springer, Heidelberg (2012)
14. IEEE 1014-1987, VME Specification
15. ANSI/VITA 1-1994; VME64 Specification
16. Bartys, M., Zbrzezny, L., Antoniewicz, A., Putz, B.: Real-time Single FPGA-Based Multimodal Image Fusion System. Accepted at 2012 IEEE International Conference on Imaging Systems and Techniques, IST 2012 (2012)

# Registration of Infrared and Visible Images Based on Edge Extraction and Phase Correlation Approaches

Jan Klimaszewski, Marcin Kondej, Mariusz Kawecki, and Barbara Putz

Warsaw University of Technology, Institute of Automatic Control and Robotics,
ul. św. Andrzeja Boboli 8, 02-525 Warsaw, Poland
janklimaszewski@wp.pl, marcin@kondej.eu, mariusz.kawecki@neostrada.pl,
bputz@mchtr.pw.edu.pl

**Summary.** The paper presents test results of several edge detectors intended to be used for multimodal image registration. It was found that application of edge detectors before running phase correlation algorithms allows achievement of good quality of registration and assure simultaneously high performance of registration necessary particularly in real time applications. Canny edge detector implemented in OpenCV and its simplification, Harris edge detector and two types of morphological edge detectors: basic and extended have been tested. It has been shown that OpenCV implementation of Canny algorithm gives best translation coefficients for IR and TV images registration. The results show that edge extraction with subsequent phase correlation is a very promising for multisensor image registration.[1]

## 1 Introduction

Registration of infrared and visible images is a typical problem to be solved in many supervisory and monitoring systems, which provide real time fusion of IR and TV images. The high quality fused image combines the complementary information from each sensor. The quality of image alignment is a key point of overall fusion process. Therefore, this alignment is still a serious, challenging problem particularly when thinking about real-time applications.

The high performance of image registration can be achieved by application of fast phase correlation method [1] due to employing global solutions using Fast Fourier Transform. This approach however, gives good quality results mainly in case of single-modal image registration. There are several problems with multisensor images, namely: the images differ in quality (e.g. IR are typically low resolution and noisy images) and some features in one image may not be observed in other. For multimodal images the mutual information methods [2] are popular and especially suitable for medical image registration,

[1] This work has been supported in part by the research project No ăO ăR00ă 0019 ă07 of National Centre for Research and Development in Poland.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 153–162.
springerlink.com 

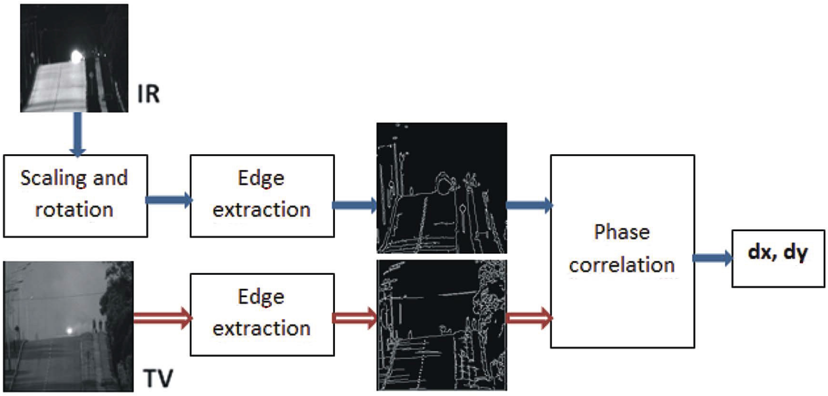



**Fig. 1** The schematics of proposed image registration process

but their computational complexity is huge. The feature based methods are preferred [2, 3, 4, 5, 6, 7] for IR and TV images registration, but none of them can be considered as very reliable and very efficient simultaneously.

Based on good results of phase correlation method applied for registration of IR and TV images, we decided to include additionally edge extraction method in a pre-processing phase. For simplicity let assume that scale and rotation be constant during image registration and only the shift calculation needs to be performed online (Fig. 1). Several edge detectors have been tested in order to solve the problem. After short presentation of techniques and edge detectors used (section 2-3), the test procedure (section 4) and comparison of results (section 5) with subsequent conclusions are further presented.

## 2 Phase Correlation

Assume that TV and IR images are coarsely registered. Fine registering of these images in terms of translation offsets might be obtained based on phase correlation method [1]. Firstly, cross-power function R called sometimes as cross-phase spectrum is calculated:

$$\mathrm{R} = \frac{I_{IR}\overline{I_{TV}}}{\left|I_{IR}\overline{I_{TV}}\right|} \tag{1}$$

In formula (1) Fourier transforms are denoted as: $I_{IR}$ and $I_{TV}$ of IR and TV images and conjugate Fourier transform $\overline{I_{TV}}$ . Next, very important is a setting of proper filtration of the R in order to achieve stable and univocal results of phase correlation approach. Finally, the inverse Fourier transform $F^{-1}(\mathrm{R})$ takes place. The shift values $(x_{\max}, y_{\max})$ for registering of IR and TV images are obtained as coordinates of the peak of cross-correlation function $F^{-1}(\mathrm{R})$:

$$x_{\max}, y_{\max} = \arg\max_{x,y} F^{-1}(\mathrm{R}). \tag{2}$$

## 3 Edge Detectors

Below we present all kinds of detectors used in the test procedure for feature-based stage.

### 3.1 Canny Edge Detector

Two implementations of original Canny edge detector [8] have been tested: OpenCV Canny implementation and the simplified one. OpenCV Canny implementation meets demands of classical Canny algorithm:

1. denoising of the source image $I(x, y)$ by means of Gaussian filter
2. calculation of gradient images $G_x = \partial I_x/\partial I$ and $G_y = \partial I_y/\partial I$ using the Sobel filters
3. detection of one of four directions of edges transfers: 0, 45, 90 or 135 degrees: $\alpha = \arctan(G_x/G_y)$
4. calculation of local maxima of $G = \sqrt{G_x^2 + G_y^2}$ for four possible angles $\alpha$
5. if for any point there is a maximum of the value of G and this value is greater than some threshold value, then this point is being assumed to be the edge point; threshold is realized with hysteresis to obtain one-pixel wide lines representing edges of the original image
6. iterative elongating and connecting the edges to obtain continuity of edges, as well as deleting the singular points previously qualified as edges.

In the simplified Canny implementation, in order to achieve real time implementation, the step No 6 has been omitted.

### 3.2 Harris Edge Detector

To test other edge extraction methods for multimodal image registration we used Harris corner detector [9, 10] being implemented in OpenCV library. This detector calculates autocorrelation matrix M, filled with gradient covariation values of some neighbourhood of every input image pixel. The obtained output image is the result of calculating:

$$\det(M) - k * \mathrm{trace}(M)^2 \tag{3}$$

where $k$ is referred as Harris detector free parameter. To get Harris corners, local maxima of this output image should be calculated. But we omit this step to preserve more dense image representation.

### 3.3 Morphological Edge Detectors

Morphological edge detector based on grey scale operations: dilation and erosion was tested. There were defined max and min operations of an image $I(x, y)$ and structuring element $ES$. Structuring element $ES$ denotes two-dimensional regular subset of image, typically being a square $3 \times 3$ or $5 \times 5$ etc. with constant intensity and with origin defined, usually in the middle of the square.

Grey level dilation is defined as transformation, which assigns a maximum value of the result of operating on intensities $I(m, n)$ of points covered by structuring element $ES$, for every pixel $(x, y)$ of the image when the origin of this element is placed at pixel $(x, y)$:

$$I_D(x, y) = \max_{m,n \in ES} I(m, n) \circ ES \tag{4}$$

Analogously, we define grey level erosion:

$$I_E(x, y) = \max_{m,n \in ES} I(m, n) \circ ES \tag{5}$$

The "$\circ$" denotes any operation on two arguments: intensity of pixels of the image and intensity of pixels of structuring element. Particularly, operation "$\circ$" can be empty or can be arithmetic sum "+", but generally it can be more complex.

To extract edges of the image, i.e. to emphasize the boundaries between regions, we can use the basic definition of morphological gradient [11]:

$$G_1(x, y) = I_D(x, y) - I_E(x, y) \tag{6}$$

Two other definitions (usually denoted as internal and external gradient, respectively) are also used:

$$G_2(x, y) = I(x, y) - I_E(x, y) \tag{7}$$

and

$$G_3(x, y) = I_D(x, y) - I(x, y) \tag{8}$$

All these three kinds of morphological gradient have been used as basic morphological edge detector (with three different parameters) during research tests.

We have tested also an extended type of morphological gradient, with image histogram equalization before gradient calculating. It brought significant results in noise sensitivity reduction.

Different modifications of morphological edge detector are also possible, using more general definitions of "$\circ$" operation or by changing the shape or intensities of structuring element. Some types of morphological gradient modifications can be found in [12, 13, 14], other are currently being developed for our tests and application.

# 4 Test Procedure

We have tested registration procedure for 18 different image sequences and for 5 edge detectors presented above: Canny detector in OpenCV implementation (marked as CannyCV), simplified Canny detector, Harris detector in OpenCV implementation (marked as HarrisCV), basic morphological gradient and extended (modified) morphological gradient.

Test procedure used algorithm with four different sets of parameters. Each set of parameters was tested by means of the same sequence of input images. During test procedure we checked whether search range ±32 or ±64 pixels is more appropriate. Simultaneously we tested also on-line smoothing filter.

Fig. 2-4 show examples of registration results (values of translation in directions X and Y). Vertical axis represents registration values and horizontal axis represents ordinal number of image pair from images set.

## 4.1 Test Results without Smoothing Filter

Fig. 2 shows results got from image registration algorithm in a search range of ±32 pixels without application of any smoothing filter. One can see that algorithm does'nt seem to find good registration values. The error computed as a sum of mean values of absolute differences between reference and calculated values for X and Y axis equals 75 pixels. Fig. 3 shows results when searching in a range ±64 pixels; tx and ty values are very close to reference registration values groundtx and groundty. Error calculated as before equals 8.5 pixels in this case.

## 4.2 On-Line Smoothing Filter

We applied an on-line smoothing filter in time domain in order to achieve overall error reduction. The main idea was to obtain the most recent value but simultaneously throw out values that are likely to be effects of noise or temporary algorithm failures. For example, we wouldn't like to have values with very high gradient value because it's very unlikely that they are valid for registration system (Fig. 3).

Results achieved with application of on-line smoothing filter turned on are shown on Fig. 4. Smoothing filter allows reducing the error for this exemplary image sequence to 6.6 pixels.

The filter can be summarized as follows:

1. Calculate histogram out of recent registration values.
2. For each of recent registration values:
   a) calculate its score accordingly to equation:

```
score = score_old * score_rare,
and:
score_old = k_old^j
```

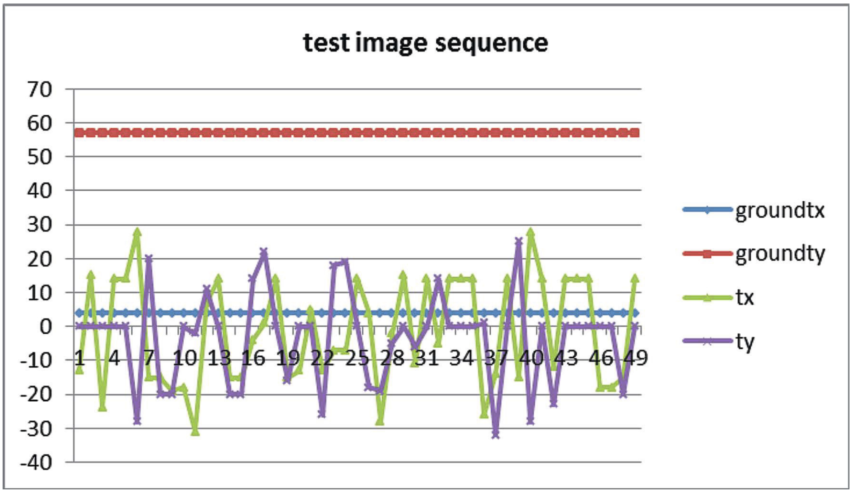



**Fig. 2** The results of aligning IR and TV images; algorithm used: Canny simplified, in range ±32 pixels; tx, ty - calculated values of translation in directions X and Y; groundtx, groundty - reference values of translation


**Fig. 3** The results of aligning IR and TV images; algorithm used: Canny simplified, in range ±64 pixels; tx, ty - calculated values of translation in directions X and Y; groundtx, groundty - ideal values of translation


**Fig. 4** The results of aligning IR and TV images, algorithm used: Canny simplified, in range ±64 pixels with smoothing filter; tx, ty - calculated values of translation in directions X and Y; groundtx, groundty - reference values of translation

```
score_rare = k_rare^(max(histogram)
 -histogram(actual_registration_value)),
```

where:
`k_old`, `k_rare` - free parameters of filter, should be in the range $< 0, 1 >$, `j` - index of value in recent values buffer, for example 0 for actual value, 1 for previous value etc.

b) output the value with greatest score value.

To sum up, filter puts `k_old` penalty for value not being actual and `k_rare` penalty for value not being frequent enough. Tuning up these parameters gives different filtration results. In our work we have got best experimental results with `k_old` $= 0.95$ and `k_rare` $= 0.6$. Considering that buffer of recent values is initialized with zeros, it would explain first few registration values shown on Fig. 4 to be all zeros.

## 5 The Comparison of Results

In Table 1 we present the results of aligning IR and TV images for all 18 synchronized, multi various tested sequences, in range ±64 pixels and with smoothing filter, for 5 edge detectors. For every sequence we have checked the best edge detector (1st winner method, i.e. giving the least maximal error) and the 2nd winner method.

Different detectors, quite often morphological, have been found the winners for different sequences, however most often it was the method no 1, i.e. Canny

**Table 1** The comparison of results for two best edge detectors and results of registration without edge extraction; the enumeration of detectors is given in legend

| sequence | 1st winner method | 2nd winner method | max error for 1st winner | max error for meth. no 1 | max error for phase corr. only |
|---|---|---|---|---|---|
| 1 | 2 | 1 | 2.02 | 2.11 | 39.11 |
| 2 | 4 | 2 | 0.72 | 1.74 | 1.40 |
| 3 | 1 | 5 | 1.15 | 1.15 | 16.13 |
| 4 | 3 | 1 | 0.30 | 0.98 | 1.26 |
| 5 | 0 | 1 | 0.60 | 1.55 | 0.60 |
| 6 | 1 | 2 | 0.74 | 0.74 | 52.49 |
| 7 | 2 | 4 | 1.06 | 3.40 | 3.64 |
| 8 | 5 | 1 | 1.26 | 1.38 | 1.53 |
| 9 | 5 | 4 | 2.32 | 32.70 | 8.55 |
| 10 | 1 | 4 | 0.57 | 0.57 | 1.47 |
| 11 | 1 | 5 | 1.87 | 1.87 | 2.66 |
| 12 | 5 | 1 | 2.53 | 3.06 | 3.06 |
| 13 | 3 | 0 | 1.79 | 1.83 | 1.79 |
| 14 | 1 | 4 | 1.74 | 1.74 | 3.23 |
| 15 | 3 | 0 | 0.30 | 0.32 | 0.30 |
| 16 | 1 | 0 | 4.34 | 4.34 | 5.00 |
| 17 | 1 | 3 | 4.94 | 4.94 | 12.60 |
| 18 | 0 | 1 | 3.15 | 3.36 | 3.15 |
| most often winner | 1 | 1 | | | |
| most often 2nd winner | 3 and 5 | 4 | | | |

| method | |
|---|---|
| 0 | phase correlation |
| 1 | CannyCV+ph.corr. |
| 2 | HartisCV+ph.corr. |
| 3 | basic morph.+ph.corr. |
| 4 | Canny simple+ph.corr. |
| 4 | modif.morph+ph.corr. |

detector in OpenCV implementation. Therefore for every sequence we compare the error with the use of this detector, the error with the use of the best detector for analyzed sequence and the error with the use of phase correlation solely. Exemplary pair of IR-TV images for sequence no 1, with corresponding pairs of edges extracted by methods 1, 2 and 3 are presented in Fig. 5.

**Fig. 5** Exemplary sequence (no 1): a, e: TV-IR images; b, f: CannyCV edges for TV-IR; c, g: HarrisCV edges for TV-IR; d, h: basic morphological edges for TV-IR

Assuming that registration with error more than 5 pixels should be treated as failed, we can see that registration using phase correlation failed 5 times (of all 18) cases. The preliminary edge extraction phase improved the results in 4 cases. The sequence No 9 is the only case where we have got total failure. But it was the specific image of the sky, without any object on it, and the edges of poorly visible clouds were detected only in IR image.

## Conclusions

The presented results, obtained for varied sequences typical in monitoring systems, indicate that we can found really good solution for efficient IR-TV images registration (assuming translations only) using the feature-based approach: edges extraction by Canny detector with subsequent phase correlation method. These results can be useful for real-time image fusion monitoring system and simultaneously for further research on efficient multisensor image registration.

## References

1. Reddy, B.S., Chatterji, B.N.: An FFT-based technique for translation, rotation and scale-invariant image registration. IEEE Transactions on Image Processing 5, 1266–1271 (1996)
2. Zitová, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing 21(11), 977–1000 (2003)

3. Sheng, Y., Yang, X., McReynolds, D., Valin, P., Sevigny, L.: Visible/IR Battlefield Image Registration using Local Hausdorff Distances. In: FUSION 1999, July 6-8, vol. 2, pp. 803–810 (1999)
4. Heather, J.P., Smith, M.I.: Multimodal Image Registration with Applications to Image Fusion. In: 7th International Conference on Information Fusion, vol. 1, pp. 1–8 (2005)
5. Orchard, J.: Globally optimal multimodal rigid registration: an analytic solution using edge information. In: IEEE Conf. on Image Proc., pp. 485–488 (2007)
6. Heather, J.P., Smith, M.I., Sadler, J., Hickman, D.: Issues and challenges in the development of a commercialised image fusion system. In: Proc. SPIE, pp. 1–12 (2010)
7. Wang, B., Wu, D., Xu, W., et al.: A New Image Registration Method for Infrared nad Visible Images. In: 2010 3nd International Congress on Image and Signal Processing, CISP 2010, pp. 1745–1749 (2010)
8. Canny, J.: A Computational Approach To Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligence 8(6), 679–698 (1986)
9. Harris, C., Stephens, M.J.: A combined corner and edge detector. In: 4th Alvey Vision Conference, pp. 147–152 (1988)
10. Derpanis, K.G.: The Harris Corner Detector. Computer and In. Science 2, 1–2 (2004)
11. Lee, J., Haralick, R.M., Shapiro, L.G.: Morphologic edge detection. IEEE Journal of Robotics and Automation RA-3(2), 142–156 (1987)
12. Qu, G., Wood, S.: Performance of a modified Gray Level Morphological Gradient with Low Sensitivity to Treshold Values and Noise. In: Conf. on Signals, Systems and Computers, pp. 931–935 (2000)
13. Zhang, L.D.: An improved Morphological Gradient Edge Detection Algorithm. In: Proceedings of ISCIT, pp. 1233–1236 (2005)
14. NagaRaju, C., NagaMani, S., Prasad, G., Sunitha, S.: Morphological Edge Detection Algorithm Based on Multi-Structure Elements of Different Directions. IJICIT Journal 1(1), 37–43 (2011)

# FPGA Implementation of Decomposition Methods for Real-Time Image Fusion

Adrian Antoniewicz

Warsaw University of Technology, Institute of Automatic Control and Robotics,
A. Boboli 8, 02-525 Warsaw, Poland
aantoniew@gmail.com

**Summary.** The paper presents an efficient FPGA implementation of decomposition methods intended for real-time fusion. FABEMD (Fast and Adaptive Bidimensional Empirical Mode Decomposition) and Laplacian pyramid methods[1] were examined. The real-time image processing was achieved by means of special image buffers with parallel and immediate data access and by application of paralleled and pipelined calculation flow ensuring constant processing latency.

## 1 Introduction

The aim of image fusion is to combine images from multiple sensors capturing the same scene in order to generate combined image, which contains possibly all details of input images. Resulting fused image should be easily analyzed by an operator or video system. Images being fused can be taken at different times and originate from multimodal sensors with different spectral sensitivity. To achieve acceptable results of image fusion, some parameters of images should be appropriately set.

This paper describes an efficient FPGA[2] implementation of fusion decomposition methods on example of FABEMD and Laplacian pyramid. Those methods were chosen for implementation considering discussion presented in [3]. Real-time processing of both algorithms was achieved by application of FPGA which gives possibility to perform parallel and pipelined computations. Fast input image buffer was realized to reduce the number of data memory transfers. Also other fusion algorithms based on image decomposition could be efficiently implemented when using proposed image buffer structure. Implemented algorithms, together with a low-level implementation, allow to develop a real-time video fusion system that may be used for example with visible and infrared video inputs.

[1] This work has been supported by the research project No O R00 0019 07 of National Centre for Research and Development in Poland.

[2] FPGA - Field Programmable Gate Arrays.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 163–170.
springerlink.com 

This paper is organized as follows: In Section 2 a brief description of the FABEMD algorithm is given. Section 3 describes the Laplacian pyramid algorithm. Section 4 presents the implementation details of the discussed algorithms and image buffers in FPGA. Finally, results, efficiency and performance of algorithms are discussed in Sections 5 and 6.

## 2 Brief Description of FABEMD Method

FABEMD algorithm [7] is a simplified BEMD[3] method based on image decomposition into series of zero-mean, oscillatory subsignals, so-called BIMFs[4]. In [8] the proper quality of FABEMD-based fusion was demonstrated, although not in the context of multi-modal image fusion, but rather in case of multi-focus images. FABEMD offers a significant reduction of computation time in comparison to BEMD. Fusion based on FABEMD can be summarized in the following steps:

1. Perform FABEMD decomposition of both initial images.
2. For each decomposition level combine values of two respective BIMFs.
3. Combine two residues.
4. Sum up all combined components to obtain the result of fusion.

Complete FABEMD fusion algorithm description was presented in [9].

## 3 Laplacian Pyramid Algorithm

Introduced by Burt [6] in 1983 Laplacian pyramid algorithm was a precursor of multi-resolution methods and currently is used in a wide range of applications besides image fusion. The algorithm is based on pyramid generation, which is made by image low-pass filtering and subsampling by a factor of two. The kernel of the low-pass filter is usually 5×5 window with Gaussian coefficients. As a result, the pyramid of subsampled images with reduced spectral bandwidth is obtained. Fused pyramid is calculated from input image pyramids using selection of pixel with higher intensity. This selection method helps to keep proper contrast ratio in final image. Fused output image is reconstructed from pyramid.

## 4 FPGA Implementation

Fusion algorithms described above are based on image decomposition to the form of so called subimages. Intensity of fused pixels rely on neighboring

[3] BEMD - Bidimensional Empirical Mode Decomposition.
[4] BIMF - Bidimensional Intrinsic Mode Function.

pixel intensity values and intensity of pixels in subimages. This requires to perform massive calculations and transfers to memory per each processed pixel. Therefore, a DMA channel between image buffer and VGA controller was realized in order to display fused image. In addition, external video codecs require to be appropriately configured.

FPGA technology was chosen to meet critical time requirements and high computational power demands. In addition, FPGA gives possibility to apply parallel and pipelined processing scheme what significantly reduce computation time. A structure of designed video system is shown in Fig. 1. Images from input codecs are captured into DDR memory by a FPGA components. Image frame reader transfers output image from buffer to output codec. Access to memory is available through Avalon Bus which also provides arbitration of bus transactions. Image fusion is performed by implemented FABEMD and Laplacian fusion components.

Special image buffer was provided to reduce amount of accesses to frame-buffered image in DDR memory. NIOS[2] software processor IP was used in order to process non-time-critical computations and algorithms, initialize video codec, configure designed fusion components and buffers. The structure of the designed system is an effective combination of FPGA performance and processor versatility and could be treated as a complete video processing system in a single FPGA chip.

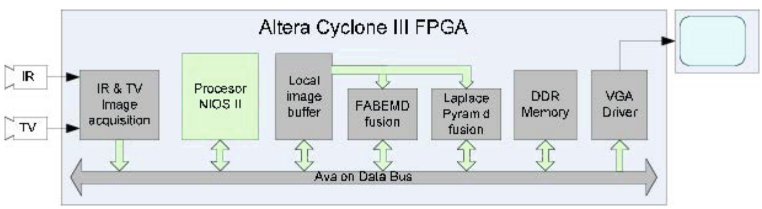



**Fig. 1** Designed FPGA video processing system structure in Altera Cyclone III Device[1]

## 4.1 Implementation of Image Buffers

Laplacian and FABEMD fusion algorithms calculate output fused pixel values using 2-D window functions of input images on each decomposition level. It means that for each processed pixel there is a need to read multiple original input pixel values (for 1st-level fusion $5\times5$ pixel image square is required, for 2nd-level fusion $9\times9$ pixel square, and for 3rd-level fusion - $17\times17$ pixel square). When images at higher resolutions are fused, there is possible to meet memory bus throughput limits and pass over critical processing times. To preserve such situations, there were realized local fast buffer with immediate

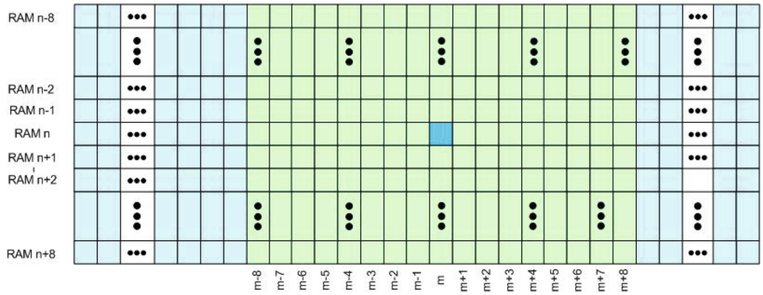



**Fig. 2** Local image buffer structure with parallel access to 18x18 pixel area

and parallel access in order to reduce number of transactions to and from DDR memory during fusion process.

Implemented local buffer reduces number of read operations from image DDR memory from hundreds per pixel (3-level fusion needs $2 * 17 * 17$ read operations) to only two per processed fused pixel. For 3-level image fusion implementation 17×17 pixel square buffers were used. Fig. 2 presents idea of realization of local image square buffer. The image buffer consists of 18 internal FPGA Block RAM, where each Block RAM memory stores one image line (typically 640 pixels) and previous RAM memory stores previous image line. Current processed pixel value read from DDR memory is written to RAM pointed by modulo 18 counter. The counter pointer is incremented at the end of processed image line. This scheme creates image line cyclic buffer.

Implemented image line cyclic buffer has parallel access to 1×18 image area (because each Block RAM could be read once per clock cycle). To achieve parallel access to 17×17 pixel area the second stage of buffer was realized. It was implemented similar to FIFO structure on 17×17 FPGA pixel registers. During processing $x_i$ pixel the data from address $i + 8$ of each Block RAM are read to group of 17 registers. During processing of pixel $x_{i+1}$, the data from previous group of registers are shifted to the next group. As a result, array of 17×17 registers with pixel intensities is achieved. Such structure of image buffer allows to independent and parallel access to each pixel of the image square at the same time and gives possibility to remarkable increase processing speed in comparison to tradition fusion scheme. In addition, other fusion algorithms based on image decomposition could be implemented using presented local buffer.

### 4.2 FABEMD Implementation

FABEMD fusion algorithm requires many calculations on input data to achieve output pixel value. Using single thread computations it is not possible to meet critical real-time requirements. Paralleled and pipelined computation

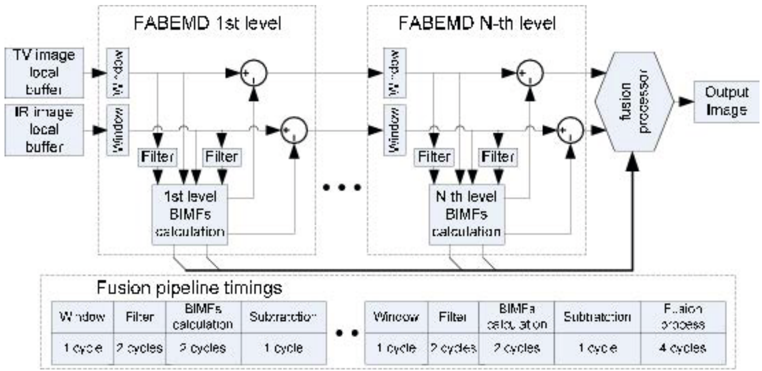



**Fig. 3** FABEMD fusion FPGA implementation processing scheme

scheme must be realized to meet all requirements. Real-time fusion process was achieved using presented in Section 4.1 fast input buffers.

Processing scheme for FABEMD fusion was presented in Fig. 3. For each decomposition level different window size was used (1st level 3×3, 2nd level 5×5, 3rd 7×7 pixels). Filter block using proper window size perform MIN/MAX filtering and smoothing function. In the next step, BIMFs are calculated by subtraction of current pixel and output values of Filter block. On each decomposition level, two respective BIMFs are calculated using maximum value select function. Filter block value is transferred to the next decomposition level. To obtain a result of fusion, an average value from last decomposition level is added to combined values achieved on each level. All computation blocks contain basic arithmetic calculations which are very easy for implementation in FPGA. Pipelined processing scheme was used to meet timing requirements (see Fig. 3). Each calculation block could be started in each clock cycle and each operation is performed in constant time, so the output latency is constant. FABEMD fusion is realized in real-time with constant latency because of usage of paralleled and pipelined scheme with fast input image buffer.

### 4.3 Implementation of Laplacian Pyramid Algorithm

Processing scheme for 3-level Laplacian pyramid fusion was presented in Fig. 4. Similar to FABEMD implementation, fast input buffers presented in Section 4.1 were used for input data. Laplacian pyramid algorithm is divided into two parts: Laplacian pyramid calculation and image synthesis [5]. Value of Laplacian pyramid for each pixel is calculated as a difference between image on current level and Gaussian filter result. Coefficients presented in Table 1 are used for Gauss filtering function implementation. Filter coefficients were normalized to 8 bit fixed point values. This reduces significantly processing

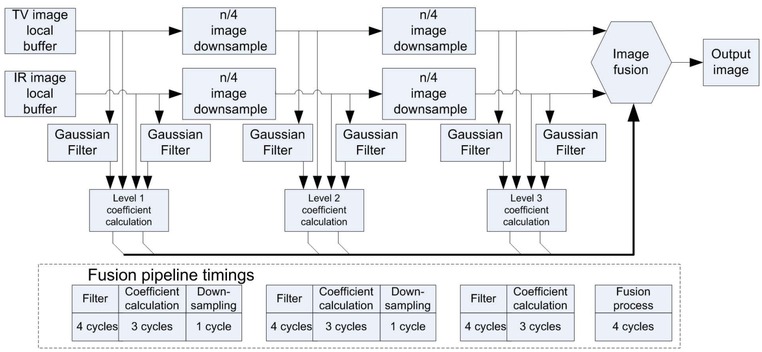



**Fig. 4** Laplacian pyramid fusion FPGA implementation processing scheme

**Table 1** Gauss filter coefficients

| | n-2 | n-1 | n | n+1 | n+2 |
|---|---|---|---|---|---|
| m-2 | 1 | 4 | 6 | 4 | 1 |
| m-1 | 4 | 16 | 24 | 16 | 4 |
| m | 6 | 24 | 36 | 24 | 6 |
| m+1 | 4 | 16 | 24 | 16 | 4 |
| m+2 | 1 | 4 | 6 | 4 | 1 |

time. Fixed point calculations give also possibility to replace multiplications by sum of shifted values ($x_i * 36 = x_i * 32 + x_i * 4 = x_i << 5 + x_i << 2$). Downsampling is performed between each decomposition level. It is done by pointing out each second pixel of current line using simple modulo 2 counter.

Similarly to FABEMD fusion implementation, all computation blocks perform basic arithmetic calculations. Pipelined processing scheme with constant latency was used to meet timing requirements (see Fig. 4).

## 5 Implementation Results

Presented fusion algorithms were developed on Altera Cyclone III Video Kit[1]. The structures of computing blocks and components were designed in order to reduce number of combinational logic and to increase maximum clock frequency. FABEMD and Laplacian pyramid fusion algorithms were tested with main clock frequency equal 166MHz. Operating frequency was limited by specification of DDR memory chips installed in development Kit. High maximum frequency give possibility to obtain fusion of high resolution images.

**Table 2** FPGA resource requirements for fusion modules

| Resource type | Input buffer | Fabemd used | Laplacian used | Total available |
|---|---|---|---|---|
| Logic cells | 7844 | 10016 | 9516 | 119088 |
| Registers | 4624 | 3238 | 2928 | 119088 |
| Memory Bits | 617472 | 0 | 0 | 3981312 |
| DSP elements | 0 | 0 | 0 | 576 |

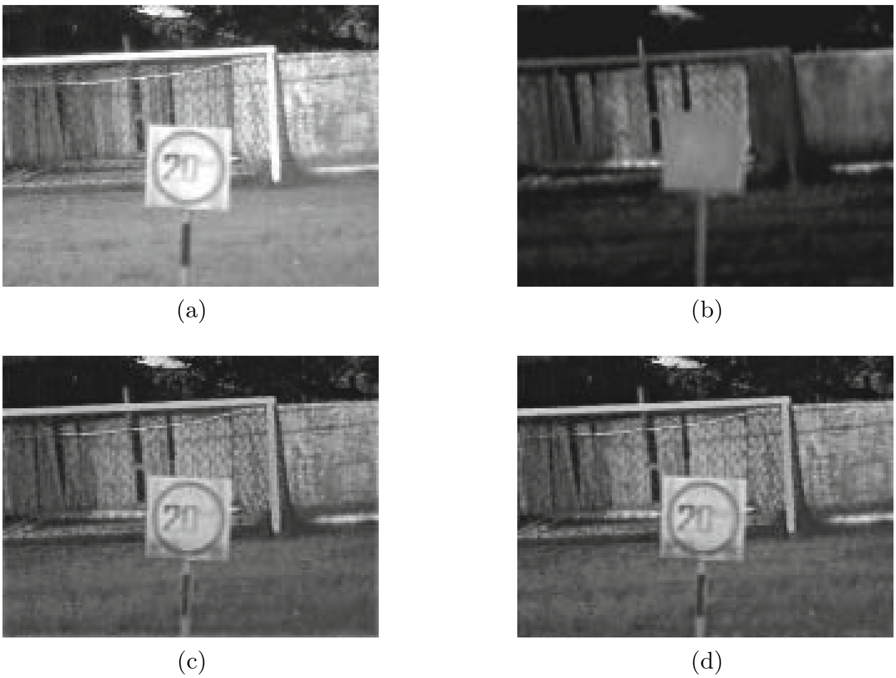



**Fig. 5** Fusion example (a): VISible camera; (b): InfraRed camera; (c): FABEMD fusion results; (d): Laplacian pyramid fusion results

The resource requirements for designed system are given in Table 2. In comparison with other multi-level fusion implementation [4] it requires small amount of RAM and do not need to use DSP elements.

Image fusion results are shown in Fig. 5. Please note that IR and TV image observe the same scene but different details are visible. For example a road sign is visible only in TV image. After fusion process, road sign, football goal and background are visible. Fusion result for FABEMD method in this case is less blurred but in other conditions Laplacian method give better results.

## 6 Conclusions

In this paper efficient FPGA implementation of decomposition method for image fusion in FPGA were presented. Proposed system architecture with special image buffering scheme, paralleled and pipelined computation scheme give possibility to perform real-time fusion of video streams. Developed prototype of real-time fusion system is able to process 25 pairs of images frames per second with resolution of 640×480 pixels. Considering typical low resolution of IR camera, this results are highly satisfactory. Presented system architecture is very flexible and modular. NIOS software processor give possibility to configure the system and control fusion process using standard C program. The system can be also extended by adding new components.

**Acknowledgement.** Author acknowledges Prof. B. Putz and Dr. M. Bartyś for the assistance in project realization.

## References

1. Altera, Cyclone III Device Handbook vol. 1,2 (2010), http://www.altera.com
2. Altera, NIOS II Processor Reference Handbook (2010), http://www.altera.com
3. Putz, B., Bartys, M., Antoniewicz, A., Klimaszewski, J., Kondej, M., Wielgus, M.: Real-Time Image Fusion Monitoring System: Problems and Solutions. In: Choraś, R.S. (ed.) Image Processing & Communications Challenges 4. AISC, vol. 184, pp. 147–156. Springer, Heidelberg (2012)
4. Sims, O., Irvine, J.: An FPGA implementation of pattern-selective pyramidal image fusion. In: Proceedings of FPL 2006, pp. 1–4 (2006)
5. Antoniewicz, A., Jamrozik, W., Kondej, M., Putz, B.: Sprzetowa realizacja fuzji obrazow metoda piramidy Laplace'a w systemach nadzoru i diagnostyki. Pomiary, Automatyka, Kontrola 07, 789–793 (2011)
6. Burt, P.J., Adelson, E.H.: The Laplacian Pyramid as a Compact Image Code. IEEE Transatctions on Communications 31, 532–540 (1983)
7. Bhuiyan, S.M.A., Adhami, R.R., Khan, J.F.: A novel approach of fast and adaptive bidimensional empirical mode decomposition. In: ICEIT, pp. V3:358–V3:363 (2008)
8. Bhuiyan, S.M.A., Adhami, R.R., Khan, J.F.: Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation. EURASIP J. Adv. Signal Proc., ID728356, 1–18 (2008)
9. Wielgus, M., Antoniewicz, A., Bartys, M., Putz, B.: Fast and Adaptive Bidimensional Empirical Mode Decomposition for the Real-time Video Fusion. Submitted to Fusion 2012, Singapore (2012)

# Binary Image Comparison with Use of Tree-Based Approach

Bartłomiej Zieliński and Marcin Iwanowski

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

**Summary.** In this paper, we propose a tree-based approach to represent and compare binary images. Upon the images trees are created. The key observation is that from similar images similar trees are produced. On the other hand, upon dissimilar images unlike trees are constructed. Moreover, the degree of dissimilarity between images is proportional to the degree of dissimilarity between the trees. Hence, it is possible to express the difference between two binary images as the difference between the trees. The paper presents algorithms of creating and comparing trees as well as results, which confirm usefulness of the approach.

## 1 Introduction

Measures of image similarity are significant in the field of image processing. The problem is to determine a value which characterizes how two images are alike. Usually, the goal is to design such a measure which would correspond with human visual system. There are a lot of possible applications for image similarity measures. The first obvious one is image retrieval, which consists in finding images in a dataset. In addition, apart from being tools for image classification, target detection and object recognition, such measures are also useful for evaluation of various image processing algorithms e.g. image coding, halftoning, compression, restoration, denoising, segmentation or image registration.

Many image comparison algorithms have been proposed so far. Existing similarity indices can be roughly divided into two groups. The first contains measures which assume that compared images are of the same size. Comparison is performed pixel-wise, so that similarity value is based on relation between pairs of corresponding pixels. For example, mean square error can be calculated this way. Pixel-wise methods are somewhat naive, yet usually fast, because no initial steps are required and pixel correspondence follow pixel location in an image. There are methods which initially determine the pixel correspondence, with one-to-many relation possible. This is an important advance, because images of different proportions can be compared. However,

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 171–177.
springerlink.com 

determining the correspondence is often a time-consuming task. Exemplary methods which use pixel correspondence in calculating binary shape similarity are presented in [5] and [6].

Another way is to extract and compare features of images, like shape, edges, texture or histogram. The problem of shape similarity is addressed in e.g. [17] and [2]. In [11], the inner-distance is introduced and taken advantage of. The inner-distance is defined as the length of the shortest path within the shape boundary to classify shape images. Some methods make use of curve alignment techniques [15]. In [10], curve correspondence is applied to distinguish between shapes. In [12], the problem is approached with so called Curvature Scale Space. There are number of algorithms which base on Hausdorff distance [7] to compare binary images. These algorithms generally calculate inter-pixel distances between images. The original solution is sensitive to noise and some research [3], [13] was aimed at reducing the shortcoming. An adaptive measure of local Hausdorff distances between images is proposed in [4]. Other methods operate on edge images [14] or contours of objects [8]. With this approach, a significant problem can occur. From noised and—especially—blurred or sharpened images detected edges can differ a lot. Consequently, indices basing on edges can produce outcomes inconsistent with ones given by human observer. Outcome edge image (which is an input to the main algorithms) is highly dependent on image modifications.

In this paper, a new image dissimilarity approach is presented. We propose a method of image tree representation as well as a way to determine a dissimilarity between images. The idea consists in creating trees upon images. Then, instead of comparing images themselves, the dissimilarity value is a result of tree comparison.

The paper is organized as follows. Section 2 presents the way of creating a tree of binary image. Section 3 describes the algorithm of comparing trees. In Section 4 experimental results are presented. In Section 5 we provide paper conclusions.

## 2 Binary Image Tree Representation

The key idea of the approach is to represent a binary image as a tree. A tree $T$ is understood as a connected graph without cycles. Trees considered in the paper are rooted ones, which means that there exists one vertex in the tree, designated as a root. Parent of a vertex is the vertex connected to it on the path to the root. $|T|$ signifies the order (number of vertices of $T$).

The algorithm works as follows. In successive steps, subsets of the input set of pixels are created. Until the desired tree level is not reached, sets are recursively determined and remain in parent–child relationship. For all the sets, *characteristic elements* are calculated. The tree is constructed by connecting the characteristic elements. As they are bound with sets of pixels, the elements remain in parent–child relationship as well and a tree arises by linking the elements which fulfill the relation. This way a tree arises. In order to apply the algorithm, one has to provide the input set, a function which determines the characteristic element, a criterion which allows constructing subsets upon a set and the height of the output tree.

A binary image can be perceived as a rectangular matrix of binary scalars. We define $S = \{p\}$, where $p$ stands for pixel coordinates, as a set confined only to pixels of the same value. For binary images which depict objects on a background, it is advisable that $S$ contains coordinates of foreground object pixels.

We define a function which returns a characteristic element of a set. In our case, $f(S) = \text{mean}\big(\{p\}\big)$ of all $p \in S$. It means that the function computes coordinates of the centroid—understood as the first moment of area—of a shape defined by indices of pixels of $S$. The coordinates are bound with corresponding vertices of $T$. This way, every centroid is represented by a tree node.

A criterion $c$ is defined as follows. Let $\rho(o_1, o_2)$ be Euclidean distance between points $o_1$ and $o_2$. Let $m = \text{median}\big(\{\rho(p, f(S))\}\big)$ for all pixels $p \in S$. In other words, $m$ stands for a median distance between pixels in $S$ and its centroid. Then, $c(S) = (S_1, S_2)$, where $S_1 = \{p_i\} \subseteq S$, such that $\rho(p_i, f(S)) \leq m$ and $S_2 = \{p_i\} \subseteq S$, such that $\rho(p_i, f(S)) \geq m$. In this way, $S_1$ contains pixels closer to the centroid (or equally distant) than $m$ and $S_2$ contains pixels farther from the centroid (or equally distant) than $m$. Taking advantage of median ensures similar cardinality of subsets. As there are two subsets created in a step of our procedure, output tree $T$ is a perfect binary one.

The algorithm successively applies function $f$ to determine characteristic points for consecutive sets extracted with use of criterion $c$. The process stops when the tree is filled to the desired level. Fig. 1 illustrates the process of constructing a tree.

In different images, depicting different content, criterion $c$ will produce unlike subsets and therefore corresponding distances among centroids will differ, as illustrated in Fig. 2. On the other hand, centroids of corresponding nodes of trees created upon similar images will remain in similar configuration, due to similar subsets returned by criterion $c$. On the basis of this observation, we assume that the more similar images are, the more similar are trees. Besides, we go one step further. We assume that the more similar trees are, the more similar have to be images upon which the trees have been created. Thus, we claim that image similarity (or dissimilarity) can be expressed by tree similarity (or dissimilarity). We can estimate the extent of dissimilarity between two trees and therefore conclude upon the degree of difference between the images.

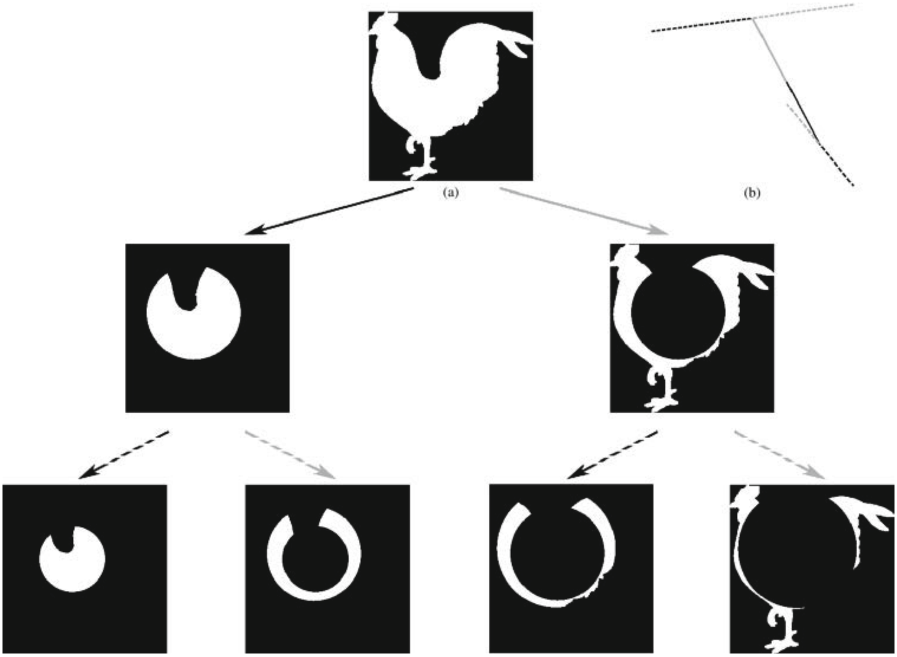



**Fig. 1** The way of creating a tree. In (a), consecutive sets returned by criterion $c$ are presented. The images are ordered as a tree, in accordance with the way the procedure works. The topmost image is the input image. Remaining images are composed of pixels which are closer or equally distant to the centroid of the parent image (left child image) and farther or equally distant to the centroid (right child image). For every set a characteristic element is calculated. The elements are bound with corresponding nodes of a tree. Every element is linked with a characteristic element of a parent node. This way, final tree (b) arises. Presented tree of height $h = 2$ is enlarged in relation to the images in (a). The color and the style of tree edges correspond to the color and the style of arrows in (a).

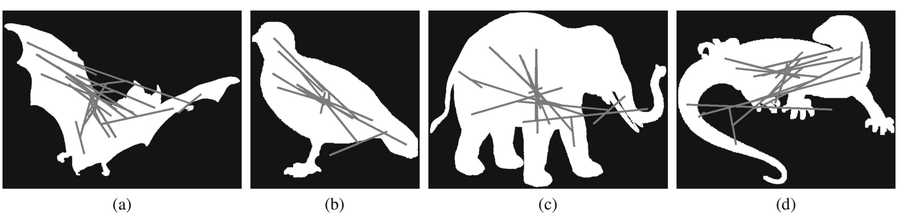



**Fig. 2** Different binary images imply different trees. Test images of MPEG7 database [1]: (a) bat, (b) bird, (c) elephant and (d) lizard exemplify the statement. The trees are of height $h = 5$.

## 3 Comparing Images

By comparing we mean determining a single scalar value which expresses the dissimilarity between the trees, and consequently between the images. The way of creating a tree effects strict structure of the tree. This is why, while investigating differences between two trees, nodes can be compared pairwise, as every node has its corresponding one in another tree (this obviously holds for trees of the same height). The rule makes the issue of comparing trees a simple one. Difference between trees is a sum of differences between pairs of nodes. For similar images, corresponding centroids' coordinates are of similar values.

Let $T_1$ and $T_2$ be trees of height $h_1$ and $h_2$, respectively. First of all, trees are aligned. It is achieved by translating the trees, so that their roots are positioned in the origin of the coordinate system. This ensures translation invariance of the trees. Then, the difference between the trees can be computed as:

$$D = \sum_{i=1}^{l} \delta_i \tag{1}$$

where $\delta_i$ stands for the difference of vertex pair $i$, $l = 2^{h+1} - 1$ with $h \in \mathbb{N}_+$ and $h \leq \min(h_1, h_2)$. This means that vertices of $h + 1$ successive tree levels are taken into consideration. The vertices which do not possess corresponding ones in another tree, are ignored.

The difference $\delta_i$ can be viewed as a "work" needed to translate a centroid of a node $i$ to the position of centroid of corresponding node of another tree and vice versa. Term "work" is an analogy to the physical quantity, which is directly proportional to the force and the distance. The distance is naturally the Euclidean distance between the centroids of a pair of corresponding vertices. The force is proportional to the number of pixels $z_{ei}$ upon which the centroid for a vertex $i$ in tree $T_e$ is calculated. This reasoning leads to the following formula:

$$\delta_i = \left(\frac{z_{1i}}{Z_1} + \frac{z_{2i}}{Z_2}\right) d_i \tag{2}$$

where $Z_e = \sum_{i=1}^{l} z_{ei}$ and $d$ stands for the distance. Note that $Z_e$ may not necessarily equal a multiple of the total number of pixels in the object. Simple assignment $Z_e$ as the total number of pixels would cause $D$ to increase along with the number of vertices $l$.

## 4 Experimental Results

The method has been subjected to experiments with use of a dataset of binary images. We have created the dataset basing on MPEG7 shape database [1].

Apart from original images, the dataset consist of deformed shapes as well. To deform original shapes we have used binary morphological operators [16]. Our dataset is composed of 40 classes of shapes. Every class contains an original shape as well as its morphologically opened and closed variants, hence there are 120 images in the dataset.

Evaluation procedure ran as follows. Every possible pair of trees, created upon dataset images, was compared. As a result, we obtained square matrix of dissimilarity values. If a reliable dissimilarity measure was used, then a simple condition was met. That is, the values calculated for objects of the same class were small, whereas those calculated for objects from among different classes were greater. To reason about the quality of the measure, we subjected the matrix of dissimilarity values to cluster analysis. Clustering was performed with use of $k$-medoids algorithm [9]. We express the reliability of the approach as the percentage of correctly clustered images. Table 1 presents the results for different height of trees.

**Table 1** Percentage Results of Tree-Based Dissimilarity Measure

| Height | | |
|---|---|---|
| $h = 2$ | $h = 3$ | $h = 4$ |
| 96.67 | 100.00 | 100.00 |

## 5 Conclusions

In this paper, we have addressed the problem of binary image comparison. Proposed method is based on image tree representation. The approach consists in constructing trees upon images and comparing the trees, rather than comparing images themselves. The difference between images can be expressed as the difference between the trees. The method is usable and effective. To evaluate the approach we have provided experimental results, which confirm the usefulness of proposed solution.

## References

1. Mpeg7 ce shape-1 part b, http://imageprocessingplace.com/downloads_V3/root_downloads/image_databases/MPEG7_CE-Shape-1_Part_B.zip
2. Ankerst, M., Kriegel, H.P., Seidl, T.: A multistep approach for shape similarity search in image databases. IEEE Transactions on Knowledge and Data Engineering 10(6), 996–1004 (1998), doi:10.1109/69.738362

3. Baddeley, A.J.: An error metric for binary images. In: Proceedings of International Workshop on Robust Computer Vision: Quality of Vision Algorithms, pp. 59–78. Wichmann Verlag (1992)
4. Baudrier, E., Nicolier, F., Millon, G., Ruan, S.: Binary-image comparison with local-dissimilarity quantification. Pattern Recognition 41(5), 1461–1478 (2008), doi:10.1016/j.patcog.2007.07.011
5. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(4), 509–522 (2002), doi:10.1109/34.993558
6. Grigorescu, C., Petkov, N.: Distance sets for shape filters and shape recognition. IEEE Transactions on Image Processing 12(10), 1274–1286 (2003), doi:10.1109/TIP.2003.816010
7. Huttenlocher, D., Klanderman, G., Rucklidge, W.: Comparing images using the Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(9), 850–863 (1993), doi:10.1109/34.232073
8. Jia, L., Kitchen, L.: Object-based image similarity computation using inductive learning of contour-segment relations. IEEE Transactions on Image Processing 9(1), 80–87 (2000), doi:10.1109/83.817600
9. Kaufman, L., Rousseeuw, P.J.: Finding Groups in Data: An Introduction to Cluster Analysis. Probability and Statistics. Wiley–Interscience, New York (1990)
10. Latecki, L., Lakamper, R.: Shape similarity measure based on correspondence of visual parts. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1185–1190 (2000), doi:10.1109/34.879802
11. Ling, H., Jacobs, D.: Shape classification using the inner-distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(2), 286–299 (2007), doi:10.1109/TPAMI.2007.41
12. Mokhtarian, F., Abbasi, S., Kittler, J.: Efficient and robust retrieval by shape content through curvature scale space. Series on Software Engineering and Knowledge Engineering 8, 51–58 (1997)
13. Paumard, J.: Robust comparison of binary images. Pattern Recognition Letters 18(10), 1057–1063 (1997), doi:10.1016/S0167-8655(97)80002-5
14. Prieto, M., Allen, A.: A similarity metric for edge images. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(10), 1265–1273 (2003), doi:10.1109/TPAMI.2003.1233900
15. Sebastian, T., Klein, P., Kimia, B.: On aligning curves. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(1), 116–125 (2003), doi:10.1109/TPAMI.2003.1159951
16. Serra, J.: Image Analysis and Mathematical Morphology. Academic Press, London (1982)
17. Veltkamp, R.: Shape matching: similarity measures and algorithms. In: International Conference on Shape Modeling and Applications, SMI 2001, pp. 188–197 (2001), doi:10.1109/SMA.2001.923389

# Analysis of Diversity Assurance Methods for Combined Classifiers

Bartosz Krawczyk and Michał Woźniak

Department of Systems and Computer Networks,
Wroclaw University of Technology, Wybrzeze Wyspianskiego 27,
50-370 Wroclaw, Poland
{bartosz.krawczyk,michal.wozniak}@pwr.wroc.pl

**Summary.** Assuring diversity of classifiers in an ensemble plays a crucial role in the multiple classifier system design. The paper presents a comparative study of selected methods which can assure the diversity by manipulating the individual classifier inputs i.e., they train learner using subspaces of a feature set or they try to exploit local competencies of individual classifier for a given subset of feature space. This work is a starting point for developing new methods of diversity assurance embedded in a multiple classifier system design. All methods had been evaluated on the basis of computer experiments which were carried out on benchmark datasets. On the basis of received results conclusions about the usefulness of examined methods for certain types of problems were drawn.

## 1 Introduction

Multiple classifier systems (MCSs) are currently the focus of intense research [12]. This conceptual approach focuses on combining knowledge of a pool of individual classifiers. There are a number of important issues that must be taken into consideration while building the aforementioned systems, such as their topology, selection of classifiers to the ensemble and fuser design. In this work we will focus on the problem of ensuring the diversity of a classifier pool. This is an important part of MCS design, as combining similar classifiers should not contribute much to MCS being constructed, apart from increasing the computational complexity. An ideal ensemble consists of classifiers with high accuracy and high diversity i.e., mutually complementary. The crucial question is how the diversity could be assured. Proposed methods exploit several approaches for diversity assurance which, for example, can be used to minimize the possibility of coincidental failure by different classifiers in the ensemble [15]. If we consider design of the classifier ensemble and creation of individual classifiers we could see similarity between them [5]. The first step focuses on selection of features and individual classifiers. The wrong selection of feature or classifiers may have a bad influence on the overall quality of classification. The second step is also very similar in both cases. The

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 179–186.
springerlink.com 

main problem associated with this part is how to select the best method of classification from the ones available at hand. In this moment it is worth recalling Wolpert's "no free lunch theorem" [20], and emphasizing that selection of good classification method does not guarantee success but the selection of bad ones in most cases leads to the deterioration of overall quality. The same similarities go on for the third step in both approaches. We try to use prepared systems and empirically asses their quality. As we mentioned above, strategy for generating the ensemble's members must seek to improve the ensemble's diversity. To enforce it we could use varying components of the MCS:

- using different input data, e.g. we could use different partitions of data set or generate various data sets by data splitting, cross-validated committee, bagging, boosting [19], because that classifiers trained on different inputs can be complementary;
- using classifiers with different outputs;
- using classifiers with the same input and output, but trained on the basis of different models or model's versions.

In this work we focus on the first group of diversity assurance methods which are filled by the methods which exploit a local specialization of individual classifiers. An individual classifier that achieves the best results is chosen from a pool for each demarcated partition of the feature space. Its answer is treated as the system answer, for all objects included in the partition. This methodology was described by Rastrigin and Erenstein [16]. Certain proposals based on this idea assume a local specialization of particular classifiers and only search for locally optimal solutions [3, 6, 18], while other methods propose dividing the feature space and selecting (or training) a classifier for each partition [14]. The main objective of the paper is to evaluate the selected methods of diversity assurance based on input manipulation, especially those focusing on choosing subsets of available features or partitioning a feature space. The results of this work will be a starting point for developing new methods of diversity assurance embedded in the multiple classifier system design.

## 2 Description of Used Ensemble Methods

Let us shortly describe the methods which we had selected for the experimental investigations.

### 2.1 Clustering and Selection

The CS algorithm [14] proposes a quite simple concept of combined classifier learning. Firstly the feature space is partitioned using a clustering algorithm and then for each of the clusters the best individual classifier is assigned according to its local competency. It is worth emphasizing the main features of the algorithm.

The CS algorithm uses clustering algorithms to divide the feature space, a task that involves separating some subsets of elements from the learning set based on their similarity [11]. It does not take into account ability of classification for each cluster. Feature space partitioning and selection of the classifiers are carried out sequentially. A natural consequence of this is the lack of feedback between steps. In the second step it is possible to find the best classifiers for the previously defined clusters. However, it is impossible to modify the shape of the clusters to adjust to the competencies of the chosen classifiers assigned to the clusters. Thus, there is no guarantee that the model obtained from the proposed partitioning is the most effective one.

### 2.2 Adaptive Splitting and Selection

On the basis of the CS analysis the AdaSS algorithm was proposed in [10]. It fuses partitioning the feature space and assigning classifiers to each partition into one integrated process, when searching for optimal parameters for the model. The main advantage of this approach is that the training algorithm considers the shape of an area to determine the content of a classifier and, conversely, that the areas adapt to the competencies of the classifiers.

### 2.3 Random Subspace

Random Subspace method [7] randomly chooses the subsets of available features and uses them to train individual classifiers. In its basic form each of the subspaces contribute equally to the overall decision. This may be perceived as a potential weakness of this method, as through the random choice of features there is a high probability of creating a subspaces consisting of feature subsets with overall weak discriminative power. But with the usage of more sophisticated fusion block or pruning methods one may select only the relevant ones, thus increasing the overall ensemble quality.

### 2.4 Random Forest

Random forest [2] is a similar method as Random Subspace but an ensemble consists of decision trees. The method combines bagging and the random selection of features in order to construct a decision tree ensemble with controlled variation.

### 2.5 Rotation Forest

The Rotation Forest generates classifier ensembles based on feature extraction [17]. It splits feature set randomly into several subsets (their number is a parameter of the algorithm) and Principal Component Analysis (PCA) is applied to each of them. All principal components are retained in order to preserve the variability information in the data.

### 2.6 Feature Driven Space Division

Feature Driven Space Division [13] is a MCS designed for high-dimensionality datasets. A feature space is split into the much smaller, disjoint subspaces. Each of them is created by the usage of a feature selection algorithm and then is used to train a classifier. This approach is based on the Random Subspaces method. By the usage of feature selection algorithms for this task it is possible to ensure that created subspaces consist only of relevant features. Then classifiers are ranked according to their individual accuracy and an ensemble pruning at a given threshold is performed.

## 3 Experimental Investigation

The main objective of experiment was to compare different methods of ensemble design based on the input manipulation.

### 3.1 Datasets

In the following experiment 10 benchmark datasets were used [4]. As we examine the behavior of ensemble methods that assure diversity using the input manipulation, we have chosen datasets with large number of features. The details of used benchmarks are given in the Table 1.

**Table 1** Details of datasets used in the experimental investigation

| No. | Name | Objects | Features | Classes |
|---|---|---|---|---|
| 1 | Audiology | 226 | 69 | 24 |
| 2 | Colic | 368 | 22 | 2 |
| 3 | Heart Disease | 303 | 75 | 5 |
| 4 | Ionosphere | 351 | 34 | 2 |
| 5 | Madelon | 4400 | 500 | 2 |
| 6 | Musk v2 | 6598 | 168 | 2 |
| 7 | p53 Mutants | 16772 | 5409 | 2 |
| 8 | Sonar | 208 | 60 | 2 |
| 9 | SPECTF Heart | 267 | 44 | 2 |
| 10 | Splice-junction | 3190 | 61 | 3 |

### 3.2 Set-Up

All experiments were carried out in R language [19] and computer implementations of used classification methods were taken from dedicated packages built-in mentioned above software or were implemented by the authors. The Rotation Forest algorithm was taken from its WEKA implementation [8], which was run from the R level with the use of RWeka [9] package.

For testing, we used a statistical test to compare the results and judge if their differences were statistically significant. For this purpose, we applied a combined 5 x 2 cv F Test [1], where feature space divisions were run independently for each of the folds.

For AdaSS and CS we have used a pool consisting of five C4.5 decision trees, each randomly pruned at different level to assure their diversity. For the remaining methods for individual classifier we have used a fully-grown C4.5 decision trees. All ensemble methods used a majority voting scheme for classifier fusion.

AdaSS was run with five cluster centroids embedded in the optimization process. Clustering and Selection was conducted with a fuzzy $k$-means clusterings, with $k = 5$. Random Forest consisted of 40 fully grown trees. Rotation Forest used the same number of trees as Random Forest and additionally applied a PCA feature space reduction approach. Random subspace split the dataset into five subspaces. Feature Driven Space Division (FDSD) used a Fast Correlation Based Filter [21] for the feature selection process, subspaces were of size equal to the 5% of the original feature space size and a ensemble pruning threshold was set to 50% of all classifiers.

### 3.3 Results

The results of experiment for the selected benchmark datasets are presented in Tab.2. We bolded the best results for each datasets. Small numbers under a given classifier's results denote indexes of the classifiers which are statistically significantly worse than a given one.

### 3.4 Discussion

The results of analyzed method are quite similar. The best results were achieved by the AdaSS, Random Forest, and Rotation Forest. It does not surprise us because the first one is a quite sophisticated method which is able to outperform the CS algorithm, and it works well especially in the case if local competencies of the individual classifier could be exploited. The Random Forest and Rotation Forest are well-established methods in the machine learning community. They are direct descendants of the Random Subspace and they usually outperform it. Feature Driven Space Division was for most cases one of the weakest classifiers, but outperformed all other ensembles in datasets no. 5 and no. 9. This is due to the fact, that FDSD was originally developed for high dimensionality problems and in such cases delivers good results. When dealing with datasets with lower number of features it creates too weak classifiers on small number of features and therefore the overall accuracy strongly deteriorates.

As we mentioned above usually difference among classifiers were not so huge, except for datasets no. 5, no. 7 and no. 9. It means that it is worth developing new methods which on the one hand can explore a wide range

**Table 2** Results of the experiment

| No. | dataset | AdaSS[1] | CS[2] | RandForest[3] | RotForest[4] | RandSubspace[5] | FDSD[6] |
|---|---|---|---|---|---|---|---|
| 1 | Audiology | 84.45 | 80.33 | **85.80** | 83.03 | 83.03 | 81.34 |
| | | 2,4,5,6 | – | 2,6 | 2,6 | 2,6 | 2 |
| 2 | Colic | **86.00** | 83.25 | **86.00** | 85.15 | 84.40 | 83.25 |
| | | 2,5,6 | – | 2,5,6 | 2,6 | 6 | – |
| 3 | Heart Disease | 80.20 | 80.20 | 83.46 | **84.15** | 82.90 | 77.45 |
| | | 6 | 6 | 1,2,5,6 | 1,2,5,6 | 1,2,6 | – |
| 4 | Ionosphere | **92.12** | **92.12** | 90.24 | 88.70 | 89.15 | 84.45 |
| | | 4,5,6 | 4,5,6 | 4,6 | 6 | 6 | – |
| 5 | Madelon | 63.50 | 54.84 | 66.50 | 65.45 | 64,05 | **68.44** |
| | | 2 | – | 1,2 | 2 | 2 | 1,2,3,4,5 |
| 6 | Musk v2 | **79.11** | 75.08 | 78.05 | 78.05 | 77.31 | 76.90 |
| | | 2,3,4,5,6 | – | 2,6 | 2,6 | 2,6 | – |
| 7 | p53 Mutants | 54.79 | 52.11 | 60.83 | 58.15 | 52.60 | **65.74** |
| | | 2,5 | – | 1,2,4,5 | 1,2,5 | – | 1,2,3,4,5 |
| 8 | Sonar | 79.45 | 75.50 | 81.15 | **82.30** | 79.45 | 74.65 |
| | | 2,6 | – | 2,5,6 | 1,2,5,6 | 2,6 | – |
| 9 | SPECTF Heart | 79.78 | 76.45 | **84.11** | 83.23 | 79.78 | 73.02 |
| | | 2,6 | 6 | 1,2,5,6 | 1,2,5,6 | 2,6 | – |
| 10 | Spice-junction | 70.00 | 66.65 | **75.90** | 71.40 | 67.25 | 74.70 |
| | | 2,5 | – | 1,2,4,5 | 2,5 | – | 1,2,4,5 |

of possible individual classifiers as candidates to the ensemble, and on the other hand will not do it randomly but on the basis of a dedicated diversity measures.

## 4 Final Remarks

The paper presents a comparative study of selected methods for assuring diversity of a classifier pool. We chose the most popular methods that are based on the input manipulation i.e., they choose set of individual classifiers trained on the basis of subsets of available features or they use a local competencies in a given area of a feature space. Experimental investigations, conducted on carefully chosen datasets and backed-up with statistical significance test showed that there is a big potential in methods that assure the diversity in dedicated, non-random way. This confirmed that it is worth developing new methods of classifier diversity assurance which can be implemented in the step of classifier selection to a pool and this work is a starting point for this purpose.

**Acknowledgement.** This work is supported by the Polish National Science Center under a grant for the period 2011-2014.

## References


1. Alpaydin, E.: Combined 5 x 2 cv f test for comparing supervised classification learning algorithms. Neural Computation 11(8), 1885–1892 (1999)
2. Breiman, L.: Random forests. Mach. Learn. 45(1), 5–32 (2001)
3. Cordella, L.P., Foggia, P., Sansone, C., Tortorella, F., Vento, M.: A Cascaded Multiple Expert System for Verification. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 330–339. Springer, Heidelberg (2000)
4. Frank, A., Asuncion, A.: UCI machine learning repository (2010), http://archive.ics.uci.edu/ml
5. Giacinto, G.: Design multiple classifier systems. Technical Report PhD thesis, Universita Degli Studi di Salerno, Salerno, Italy (1998)
6. Giacinto, G., Roli, F., Fumera, G.: Design of effective multiple classifier systems by clustering of classifiers. In: Proceedings of 15th International Conference on Pattern Recognition, vol. 2, pp. 160–163 (2000)
7. Ho, K.T.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, 832–844 (1998)
8. Holmes, G., Donkin, A., Witten, I.H.: Weka: A machine learning workbench. In: Proceedings of Australian and New Zealand Conference on Intelligent Information Systems, pp. 357–361 (1994)
9. Hornik, K., Buchta, C., Zeileis, A.: Open-source machine learning: R meets weka. Computational Statistics 24(2), 225–232 (2009)
10. Jackowski, K., Wozniak, M.: Algorithm of designing compound recognition system on the basis of combining classifiers with simultaneous splitting feature space into competence areas. Pattern Analysis and Applications 12(4), 415–425 (2009)
11. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. 31, 264–323 (1999)
12. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1), 4–37 (2000)
13. Krawczyk, B.: Classifier Committee Based on Feature Selection Method for Obstructive Nephropathy Diagnosis. In: Katarzyniak, R., Chiu, T.-F., Hong, C.-F., Nguyen, N.T. (eds.) Semantic Methods for Knowledge Management and Communication. SCI, vol. 381, pp. 115–125. Springer, Heidelberg (2011)
14. Kuncheva, L.I.: Clustering-and-selection model for classifier combination. In: Proceedings of Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, vol. 1, pp. 185–188 (2000)
15. Partridge, D., Krzanowski, W.: Software diversity: practical statistics for its measurement and exploitation. Information and Software Technology 39(10), 707–717 (1997)
16. Rastrigin, L.A., Erenstein, R.H.: Method of Collective Recognition. Energoizdat, Moscow (1981)
17. Rodríguez, J.J., Kuncheva, L.I., Alonso, C.J.: Rotation forest: A new classifier ensemble method. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(10), 1619–1630 (2006)
18. Ruta, D., Gabrys, B.: Classifier selection for majority voting. Information Fusion 6(1), 63–81 (2005)

19. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2008)
20. Wolpert, D.H.: The supervised learning no-free-lunch theorems. In: Proc. 6th Online World Conference on Soft Computing in Industrial Applications, pp. 25–42 (2001)
21. Yu, L., Liu, H.: Feature selection for high-dimensional data: A fast correlation-based filter solution. In: Proceedings of Twentieth International Conference on Machine Learning, vol. 2, pp. 856–863 (2003)

# Weighted Feature Similarity – A Nonlinear Combination of Gradient and Phase Congruency for Full-Reference Image Quality Assessment

Krzysztof Okarma

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

**Summary.** In the paper the modified Feature Similarity metric has been discussed which is based on the nonlinear combination of two elements being the basics of the recently developed Feature Similarity metric for full-reference image quality assessment. Nevertheless, the influence of the gradient magnitude and phase congruency, used as two main elements of the metric, on the perceived quality is not necessarily equal. For this reason some experiments have been conducted in order to propose the weighting coefficients, applied as the local exponents, increasing the rank order correlation coefficients with subjective quality evaluations. The verification of the obtained results has been conducted using 5 ”state-of-the-art” benchmark databases and the obtained weighted FSIM metric's performance results are better for all of them.

## 1 Introduction

Image quality assessment is one of the most rapidly developing areas of image analysis in recent years. Considering some recent publications in this area, many newly proposed metrics can be noticed which are usually much better correlated with human perception of typical image contaminations than traditionally used Mean Squared Error (MSE) or Peak Signal to Noise Ratio (PSNR). Nevertheless, even recently proposed metrics are typically defined only for greyscale images and do not utilize colour information, so there is still an open field of research related to such algorithms [1, 2]. Some of the metrics are also delivered using some specific assumptions and shortcomings e.g. related to the assumed distortion types or the relative importance of a specified data (e.g. representing the edges) reducing their universality or decreasing the achieved correlation with subjective quality scores.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 187–194.
springerlink.com 

Automatic (objective) image quality assessment methods can be divided into three major groups depending on the availability of the information related to the original undistorted (reference) image. Most of the popular universal methods, sensitive to many types of distortions and well correlated with human perception, belong to the full-reference algorithms which require the full knowledge of the reference image in order to compare some features of the analysed image with the reference one, similarly as the pixel-based MSE and PSNR metrics.

Another group of algorithms is known as the reduced-reference methods which require only partial information about the reference image (typically a relatively small set of features being compared e.g. some of the DCT or wavelet transform coefficients). The most desired approach is the usage of the no-reference (”blind”) methods, which can be applied if original image is unknown (e.g. in video transmission over lossy channels). Nevertheless, such metrics are usually sensitive to only one or two types of distortions, so their universality is much lower.

The most typical area of applications of the image quality assessment methods is probably the development of new lossy image (and video) compression algorithms as well as new image filtering methods. Since the progress in image filtration is related mainly to colour image processing, some new colour image quality metrics are very desirable.

Nevertheless, there are also some other relevant areas of applications for image quality assessment e.g. related to mobile devices, considering also changing viewing and lighting conditions. On the other hand the quality of images affects the recognition accuracy in many image analysis and pattern recognition methods including e.g. face recognition systems, register plate’s numbers recognition for Intelligent Transportation Systems etc. Since there is a high correlation between the recognition accuracy and the image quality assessment results [3], the quality metrics can be applied also for the prediction of the recognition accuracy or comparison of some other image processing algorithms.

## 2 Development of Modern Image Quality Metrics

The ”new age” of objective image quality assessment has started in 2002 with the idea of Universal Image Quality Index proposed by Zhou Wang and Alan Bovik [4]. This metric has become a starting point for the development of the probably the most popular modern metric known as the Structural Similarity (SSIM) [5]. Both metrics are sensitive to three common types of distortions: loss of contrast, luminance distortions and the most important structural changes.

In recent years some further modifications of the SSIM metric have also been proposed e.g. Multi-Scale SSIM [6], gradient based SSIM (G-SSIM) [7], Complex-Wavelet SSIM [8], 3-Component Weighted SSIM [9] or the applications of additional saliency maps.

The idea of the image quality assessment based in the similarity has also been used in recently proposed Riesz-transform based Feature SIMilarity metric (RFSIM) [10]. This metric is based on the assumption that the most relevant image regions are close to the edges which can be detected using well-known Canny filter. Assuming strong influence of low level features on the perceived quality the Riesz transform can be applied only to the nearest neighbourhood of the detected edges. Even better correlation with subjective scores can be obtained using recently proposed Feature Similarity (FSIM) metric proposed by the same authors [11] based on quite similar assumptions, which is discussed below in a more detailed way.

Apart from the image quality metrics based on the similarity, some other approaches have also been presented recently e.g. based on the Singular Value Decomposition (SVD) [12, 13, 14, 15, 16, 17, 18] but their higher computational complexity causes their less significance, especially for real-time applications.

## 3 Correlation with Subjective Scores and Image Quality Benchmarking Databases

Considering the results of the correlation of each metric with subjective evaluations some standard test image databases are typically used. The most relevant of such currently available databases are Tampere Image Database (TID2008) [19] containing 1700 images with 17 types of distortions, Categorical Subjective Image Quality (CSIQ) database [20] with 866 images and LIVE Image Quality Database [21] with 779 distorted images. The other two less significant databases are known as Wireless Image Quality (WIQ) database [22] with greyscale images having distortions typical for wireless transmission and IRCCyN/IVC [23] containing 160 test images.

The typical approach to the verification of the usefulness of the objective metrics is the calculation of their correlation with subjective scores expressed as Mean Opinion Score (MOS) or Differential MOS (DMOS) values for each of the images included in the test image databases. For this purpose Pearson's linear correlation coefficient (PCC) is used for the prediction accuracy and Spearman's and Kendall's Rank Order Correlation Coefficients (SROCC and KROCC) for the monotonicity.

Unfortunately Pearson's correlation with MOS/DMOS values even for the most recent metrics is not always high for all databases because of the nonlinear relationship between the metrics and perceived quality. This problem is typically solved using the additional nonlinear fitting usually using the logistic function, but the optimal parameters obtained for various databases differ significantly decreasing the universality of such approach. Another possible solution for compensation of such nonlinearity is the application of the combined metric [24]. Nevertheless, since the impact of the benchmark dataset on the performance of image quality metrics may be significant, the universality of each metric should be confirmed using several available databases [25].

Analysing the procedures of the subjective scores' acquisition applied for each database, it can be noticed that most results have been obtained as the effect of the comparison of images. For this reason the application of the rank order correlation coefficients, based on the sorting of the subjective and objective scores, may be a reasonable choice without the necessity of any nonlinear mapping.

The Spearman correlation coefficient can be calculated according to the following formula

$$\mathrm{SROCC} = 1 - \frac{6 \cdot \sum_{i=1}^{N} D_i^2}{N \cdot (N^2 - 1)} \tag{1}$$

where $D$ is the difference between the positions of the subjective and objective scores in two sorted score lists and $N$ denotes the length of each list equal to the number of the assessed images in the dataset.

The Kendall correlation coefficient can be computed as the difference between two probabilities related to the same and inverse order of both lists

$$\mathrm{KROCC} = 2 \cdot \frac{P - Q}{N \cdot (N - 1)} \tag{2}$$

where $P$ denotes the number of score pairs corresponding to the same image located in the same order and $Q$ in the reverse one.

## 4 Feature Similarity Index and Proposed Modifications

The idea of the FSIM metric [11] is based on the use of the phase congruency (PC) and gradient magnitude (G) values for the calculation of the local similarity index. The overall quality index for greyscale images is defined as

$$\mathrm{FSIM} = \frac{\sum_x \sum_y S(x,y) \cdot PC_m(x,y)}{\sum_x \sum_y PC_m(x,y)} \tag{3}$$

where $PC_m(x,y) = max(PC_1(x,y), PC_2(x,y))$ denotes the higher of the two local phase congruency values calculated for the original and distorted image respectively. The local similarity value $S(x,y)$ is then computed as

$$S(x,y) = \left( \frac{2 \cdot PC_1(x,y) \cdot PC_2(x,y) + T_{PC}}{PC_1^2(x,y) + PC_2^2(x,y) + T_{PC}} \right) \cdot \left( \frac{2 \cdot G_1(x,y) \cdot G_2(x,y) + T_G}{G_1^2(x,y) + G_2^2(x,y) + T_G} \right) \tag{4}$$

where $T_{PC}$ and $T_G$ are small stabilizing constants. The gradient values can be obtained using gradient convolution filters e.g. Prewitt, Sobel or Scharr (which is recommended by the authors of the FSIM index). The detailed analysis of the phase congruency measurement is presented in the paper [26]. The colour version of the metric, denoted as FSIMc, is based only on the modification of the local similarity value in the equation (3) according to

$$S_c(x,y) = S(x,y) \cdot \left[ \frac{2 \cdot I_1(x,y) \cdot I_2(x,y) + T_{IQ}}{I_1^2(x,y) + I_2^2(x,y) + T_{IQ}} \cdot \frac{2 \cdot Q_1(x,y) \cdot Q_2(x,y) + T_{IQ}}{Q_1^2(x,y) + Q_2^2(x,y) + T_{IQ}} \right]^{\gamma} \quad (5)$$

where $I$ and $Q$ are the chrominance values in the YIQ colour model and the recommended value of the parameter $\gamma$ is 0.03. Nevertheless, the results obtained for the colour version of the metric are very similar to the ones achieved using greyscale FSIM metric.

The authors of the FSIM index have assumed the equal importance of the gradient magnitude and phase congruency parts of the metric for simplicity. Nevertheless, the optimisation of the exponents may cause better correlation with subjective scores leading to the weighted Feature Similarity. In such case a more general definition should be used instead of formula (4)

$$S(x,y) = \left( \frac{2 \cdot PC_1(x,y) \cdot PC_2(x,y) + T_{PC}}{PC_1^2(x,y) + PC_2^2(x,y) + T_{PC}} \right)^{\alpha} \cdot \left( \frac{2 \cdot G_1(x,y) \cdot G_2(x,y) + T_G}{G_1^2(x,y) + G_2^2(x,y) + T_G} \right)^{\beta} \quad (6)$$

assuming that $\alpha \neq \beta \neq 1$.

Since the choice of the exponents $\alpha$ and $\beta$ influences directly on the local similarity values, some extensive tests should be conducted in order to obtain their sub-optimal values leading to higher correlation with subjective evaluations. Further modifications towards reliable colour image quality assessment may be related to the choice of another colour space instead of YIQ with additional optimisation of the appropriate exponents replacing the $\gamma$ coefficient from the formula (5). Nevertheless, modifications related to colour analysis are not the main topic of this paper and can be considered as a promising direction of future research. The experiments conducted in this research have been limited to the choice of the $\gamma$ exponent after the optimisation of the $\alpha$ and $\beta$ coefficients without changing the colour space.

The optimisation procedure allowing the proper choice of the exponents $\alpha$ and $\beta$ has been conducted using the largest available database (TID2008), mainly due to the large number of involved observers as well as the number of images and distortion types. The obtained coefficients $\alpha = 0.01$ and $\beta = 0.05$ have also been used for the calculation of the metric using four other significant image databases used in the experiments. The performance of the obtained weighted FSIM metric has been compared with the standard FSIM using the Spearman and Kendall rank order correlation coefficients and the obtained results are presented in Table 1. Assuming the new values of the $\alpha$ and $\beta$ exponents significantly differing from 1, the use of the default value of the $\gamma = 0.03$ coefficient significantly decreases the performance of the metric, so

**Table 1** Obtained results of the Spearman and Kendall Rank Order Correlation Coefficients for various databases

| Database | Spearman (SROCC) | | Kendall (KROCC) | |
|---|---|---|---|---|
| | FSIM | WFSIM | FSIM | WFSIM |
| TID2008 | 0.8805 | **0.8955** | 0.6946 | **0.7149** |
| CSIQ | 0.9242 | **0.9252** | 0.7567 | **0.7573** |
| LIVE | 0.9634 | **0.9639** | 0.8337 | **0.8351** |
| IVC | 0.9262 | **0.9269** | 0.7564 | **0.7570** |
| WIQ | 0.8006 | **0.8059** | 0.6215 | **0.6221** |

**Table 2** Obtained results of the Spearman and Kendall Rank Order Correlation Coefficients for various databases using the colour metrics

| Database | Spearman (SROCC) | | Kendall (KROCC) | |
|---|---|---|---|---|
| | FSIMc | WFSIMc | FSIMc | WFSIMc |
| TID2008 | 0.8840 | **0.9041** | 0.6991 | **0.7257** |
| CSIQ | 0.9310 | **0.9365** | 0.7690 | **0.7764** |
| LIVE | 0.9645 | 0.9639 | 0.8363 | 0.8342 |
| IVC | 0.9293 | 0.9262 | 0.7636 | 0.7587 |

much lower exponent's value should be used. Table 2 illustrates the values of the rank order correlation coefficients obtained for standard FSIMc metric and the modified one using the sub-optimal value (optimised using TID2008 database) of the coefficients with $\gamma = 0.004$. Due to the presence of only greyscale images in the WIQ database, the calculations of colour metrics have been conducted only for TID2008, LIVE, CSIQ and IVC datasets.

## 5 Discussion of Results and Conclusions

Analysing the obtained results, the advantages of the proposed weighted FSIM index for the greyscale image quality assessment can be observed for all currently available relevant datasets in terms of Spearman and Kendall rank order correlation coefficients, which correspond to the sorting of the images with various types and amounts of distortions according to their perceived quality. The application of the modified colour index also leads to better results for two most important databases but slightly worse results for LIVE and IVC datasets can be noticed. The best results have been obtained for TID2008 database and its main reason is related to the choice of this dataset for the optimisation of the exponents.

Comparing the obtained results with some other ideas presented in various publications e.g. related to the applications of Singular Value Decomposition, a similar increase of the performance can be noticed. However, a relevant advantage of the approach proposed in the paper is the fact that it does not increase the computational complexity of the metric in a significant way, preserving also the possibility of its parallel computations e.g. utilising the advantages of modern GPGPUs architecture.

Nevertheless, the application of YIQ colour space, proposed by the authors of the FSIM index, is doubtful, so more complex research should be conducted towards possible application of some other colour spaces with Feature Similarity index. Since the reliable objective colour image quality assessment is still an open field of research, such direction of future work can be considered as one of the most relevant one in the image analysis area.

## 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.: Colour Image Quality Assessment Using the Combined Full-Reference Metric. In: Burduk, R., Kurzyński, M., Woźniak, M., Żołnierek, A., et al. (eds.) Computer Recognition Systems 4. AISC, vol. 95, pp. 287–296. Springer, Heidelberg (2011)
3. Forczmański, P., Okarma, K.: Application of the central weighted Structural Similarity index for the estimation of the face recognition accuracy. Annales UMCS - Informatica 9(1), 153–163 (2009)
4. Wang, Z., Bovik, A.: A universal image quality index. IEEE Signal Proc. Letters 9(3), 81–84 (2002)
5. 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)
6. Wang, Z., Simoncelli, E., Bovik, A.: Multi-Scale Structural Similarity for image quality assessment. In: Proc. 37th IEEE Asilomar Conf. on Signals, Systems and Computers (2003)
7. Chen, G.H., Yang, C.L., Xie, S.L.: Gradient-based structural similarity for image quality assessment. In: Proc. Int. Conf. Image Processing, pp. 2929–2932 (2006)
8. Sampat, M., Wang, Z., Gupta, S., Bovik, A., Markey, M.: Complex Wavelet Structural Similarity: A new image similarity index. IEEE Trans. Image Proc. 18(11), 2385–2401 (2009)
9. Li, C., Bovik, A.: Three-component weighted structural similarity index. In: Proceedings of SPIE - Image Quality and System Performance VI, 72420Q (2009)
10. Zhang, L., Zhang, L., Mou, X.: RFSIM: A feature based image quality assessment metric using Riesz transforms. In: Proc. Int. Conf. Image Processing, pp. 321–324 (2010)
11. 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)
12. Shnayderman, A., Gusev, A., Eskicioglu, A.: An SVD-based gray-scale image quality measure for local and global assessment. IEEE Trans. Image Proc. 15(2), 422–429 (2006)

13. Mahmoudi-Aznaveh, A., Mansouri, A., Torkamani-Azar, F., Eslami, M.: Image quality measurement besides distortion type classifying. Optical Review 16(1), 30–34 (2009)
14. Mansouri, A., Mahmoudi-Aznaveh, A., Torkamani-Azar, F., Jahanshahi, J.: Image quality assessment using the Singular Value Decomposition theorem. Optical Review 16(2), 49–53 (2009)
15. Wang, R., Cui, Y., Yuan, Y.: Image quality assessment using full-parameter Singular Value Decomposition. Optical Engineering 50(5), 057005 (2011)
16. Zhang, F., Li, J., Chen, G., Man, J.: Assessment of color video quality with Singular Value Decomposition of complex matrix. In: Proc. Int. Conf. Information Assurance and Security, pp. 103–106 (2009)
17. Narwaria, M., Lin, W.: Objective image quality assessment with Singular Value Decomposition. In: Proc. 5th Int. Workshop Video Processing and Quality Metrics for Consumer Electronics (2010)
18. Narwaria, M., Lin, W.: SVD-based quality metric for image and video using machine learning. IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics 42(2), 347–364 (2012)
19. Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., Battisti, F.: TID2008 - a database for evaluation of full-reference visual quality assessment metrics. Advances of Modern Radioelectronics 10, 30–45 (2009)
20. Larson, E., Chandler, D.: Most apparent distortion: full-reference image quality assessment and the role of strategy. Journal of Electronic Imaging 19(1), 011006 (2010)
21. Sheikh, H., Wang, Z., Cormack, L., Bovik, A.: LIVE image quality assessment database release 2 (2005), http://live.ece.utexas.edu/research/quality
22. Engelke, U., Zepernick, H., Kusuma, T.: Subjective quality assessment for wireless image communication: The Wireless Imaging Quality database. In: Proc. 5th Int. Workshop Video Processing and Quality Metrics for Consumer Electronics (2010)
23. Le Callet, P., Autrusseau, F.: Subjective quality assessment IRCCyN/IVC database (2005), http://www.irccyn.ec-nantes.fr/ivcdb/
24. Okarma, K.: Combined Full-Reference Image Quality Metric Linearly Correlated with Subjective Assessment. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010. LNCS (LNAI), vol. 6113, pp. 539–546. Springer, Heidelberg (2010)
25. Tourancheau, S., Autrusseau, F., Sazzad, Z., Horita, Y.: Impact of subjective dataset on the performance of image quality metrics. In: Proc. Int. Conf. Image Processing, pp. 365–368 (2008)
26. Liu, Z., Laganiére, R.: Phase congruence measurement for image similarity assessment. Pattern Recognition Letters 28(1), 166–172 (2007)

# Ontological Imaging [O-I] with Case Studies

Nikolaos E. Myridis

Aristotle University of Thessaloniki
nmyridis@theo.auth.gr

**Summary.** In this paper we inaugurate the shaping of a new Imaging field, that of Ontological Imaging, in order to diversify the families of images which are dedicated to convey semantic content. The images of this area can be characterized as *ontological*, in distinction to images characterized by the term 'artistic' (in a broader sense). The space of applications regarding Ontological Imaging seems to be inexhaustible. Thus we cite herein two broad areas of applications, that is: (a) *imaging of mathematical entities* and (b) *sub-visual imaging of processes*. We give also two significant case studies, regarding the previous areas: (a) imaging of real numbers and (b) imaging of mathematical constants.

## 1 Introduction

In a recent paper [6] a holistic view of the Image field was inaugurated. A subfield of the proposed model [Image Eye Diagram (IED)] refers to the Ontology of Images.

Image, in general, is an entity containing an arbitrary content. The latter may thematically vary, thus determining a variety of thematic areas in Image field. It is plausible for instance, at a first stage, to categorize images into: (a) those conveying (geometrical) figures, namely the *schematic* images, and (b) those conveying symbolic meaning, the so-called *symbolic*. We can moreover distinguish another huge area of images which may be characterized as artistic. This family contains -in the wide sense- all those images stimulating the aesthetic sensibility of an observer [7]. In this paper we introduce and determine, as a counterpart of the previous family, another area of images: the *ontological images*.

## 2 Ontological Imaging

We launch now a new perspective in the field of Image: *Ontological Imaging*. We give initially a definition of it.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 195–200.
springerlink.com 

In Ontological Imaging [O-I] the goal is not the artistic assignment of information (color, shape etc.); instead, the objective is the transmission of the semantic content of information.

Each image, of course, also conveys a kind of, deeper or trivial, obvious or indiscernible, semantic content. Thus, each image, at a higher or lower level, pertains to an image of ontological hue. As well as, each image is an artistic entity by itself. In the herein introduced area (i.e. Ontological Imaging) we concern those images which are created and which, primarily or absolutely, aim to the expression and transmission of semantic information. In this context, diverse imaging applications could be cited, as for instance spectral images [9], imaging of parameters etc. However, we should state that every image has an ontological version; indeed every image conveys a small or large amount of semantic information.

Ontological Imaging is obviously a huge, distinct area of Image Science and consequently of Image Processing, with innumerable and indispensable applications. It may also correlate to Ontological Engineering [2].

In a few words we could say that this new thematic area contributes to the transition of Image Processing to a deeper and more essential stage than that of a morphological performance and investigation; that is, to (a) the content analysis and (b) the emergence and presentation of the (semantic) essence of an image, what by evidence is the most important and difficult goal.

Moreover, at this point, we could cite a simple but characteristic model of image analysis, which however is a fundamental tool for the command and transaction of each work/process in Image Processing. We demonstrate this new approach in Fig. 1, following a block- structural constructional model.

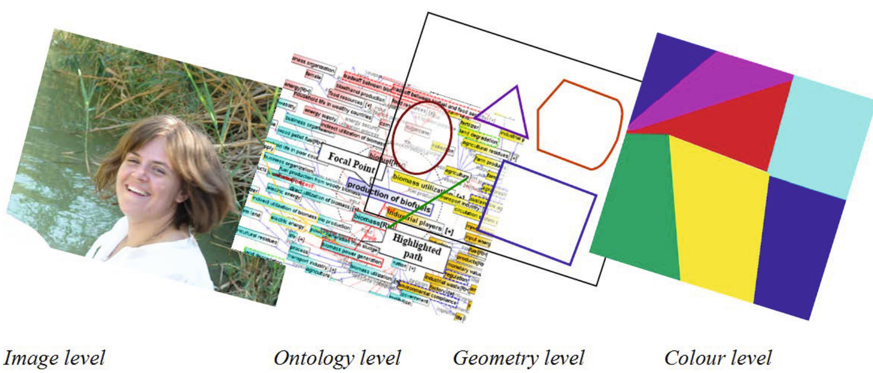



**Fig. 1** Block- structural model for images

The *Image Level* corresponds to the observable image (i.e. what we actually see). The rest three levels (*Ontology*, *Geometry* and *Colour*) are the fields wherein an image could be resolved.

*Ontology Level* refers to the semantic content of the image.
*Geometry Level* isolates the figures and forms of an image.
Finally, *Colour Level* represents the colour space of the image.

## 3 Paradigms of Ontological Imaging

To the extent of our knowledge, the starting point of O-I is this current paper. Thus, we may evidently anticipate the opening of a long -and probably inexhaustible- catalogue of prosperous applications.

For the sake of this presentation we cite here two broad and significant categories of such useful application paradigms. Case studies of these paradigms will be given in the next section.

(A) *Imaging of Mathematical Entities*. This new branch-area is of course of vital usefulness, inasmuch as it is well known that there are mathematical processes/states which can not be still imaged (!) [4].

Nevertheless, we could also propose, using O-I, many simple forms and applications of imaging of numbers, which moreover constitute useful products for the understanding, development and transmission of mathematical elements/ information.

(B) *Structural & sub-visual Imaging of (a) processes/states of the world or even (b) the whole world (Universe)*. In parallel to the difficulty for imaging of mathematical processes there is also a relative difficulty for imaging of physical states, such as, for instance, processes of structural value in the Universe. Such a case is the endeavor of combined imaging of mathematical constants in the Universe.

Common place of the preceding categories is the contribution of Ontological Imaging to the imaging of processes which were difficult (or indefinite) to be imaged earlier.

## 4 Applications of Ontological Imaging

In this section we demonstrate specific applications of Ontological Imaging. The scheme of presentation preserves the order of the O-I paradigms in the previous section.

($\alpha$) *Imaging of real numbers*. We cite at this point an alternative method for imaging of (real) number, using Ontological Imaging (Fig. 2). The index of columns represents the integer part of a number while the index of rows refers to the decimal part. The columns ($\omega$) and ($n$) indicate the order of decimal digits (magenta circles) and the number of occurrences of the corresponding decimal digit (green circles), respectively. Other relatively close applications may be found (e.g. [8]).

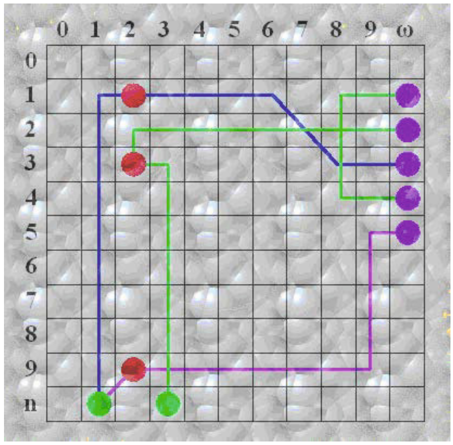



**Fig. 2** Ontological Imaging of real numbers (the number 2.33139 is depicted)

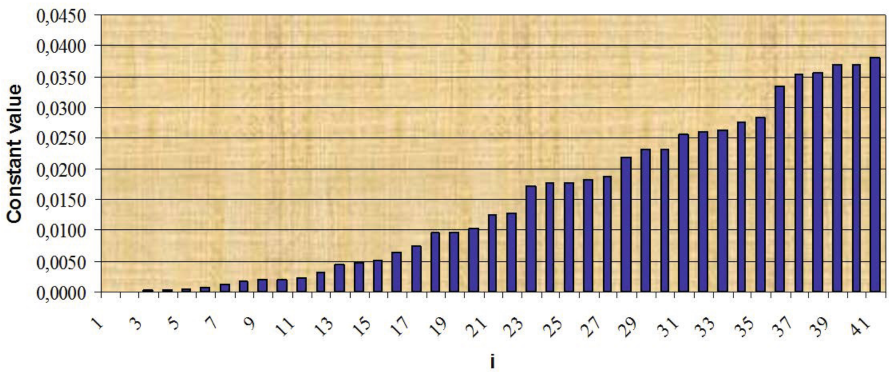



**Fig. 3** 1-D Histogram of Mathematical Constants (the window of the first 40 values is illustrated)

(b) In the second category of O-I applications we focus our interest on the problem of imaging of mathematical constants [1]. This is a structural-syntactic problem in the Universe.

We use for this purpose the mathematical constant list which can be found in [1]. The total number of constants is $n_{Mc}$=971. In the 1-D case the *Histogram of Mathematical Constants* shall be considered. Thus, a window of this histogram is illustrated in Fig. 3.

However, the 2-D case is harder. We consider as an optimal solution, the use of a continuous image field (analog image [5]) for imaging purposes. We propose an appropriate form of this field, illustrated in Fig. 4: a graph representing real numbers as a function of their decimal part. The locii of

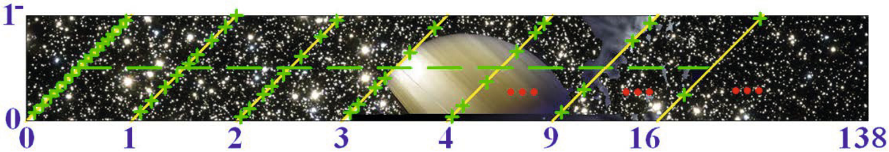



**Fig. 4** An application of O-I: *backbone of the Universe*

constants are located onto parallel line segments and the constants are marked by crosses.

Thus a form of *spondyls* (i.e. of distinct branches) is shaped, whereon the whole Universe is built. Indeed, the constructional role of the constants in the Universe is well-known. [We could also consider the symmetry axis $y = 0.5$; therefore a *backbone ( of the Universe)* results.] The max value regarding $x$-axis is $x_{\max} = 138$ (due to the max constant value (137.0359)), while $y_{max} = 1^-$ (due to the max value of the decimal part of each real number).

## 5 Conclusion

In this paper we proposed a new area of Imaging, that of Ontological Imaging. This area could also be considered as a sub-field of Ontological Engineering. O-I refers to the use of Images in order to convey semantic content. Obviously the overall space of applications spanned by Ontological Imaging is huge. We presented herein, as a starting point, two wide sub-areas of applications, those of (a) *imaging of mathematical entities* and (b) *sub-visual imaging of processes*, with case studies regarding the *imaging of (a) real numbers and (b) constants*, respectively. The latter leads to a new imaging approach offered to the endeavor for knowledge of the Universe. Finally, it would not be an overstatement that in each case, new unknown gates are opened.

## References

1. Finch, S.: Mathematical Constants. Cambridge University Press, N.Y. (2005)
2. Gomez-Perez, A., Fernandez-Lopez, M., Corcho, O.: Ontological Engineering. Springer, Berlin (2004)
3. Kozaki, K., Hirota, T., Mizoguchi, R.: Understanding an Ontology through Divergent Exploration. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part I. LNCS, vol. 6643, pp. 305–320. Springer, Heidelberg (2011)
4. Krellenstein, M.: Unsolvable Problems, Visual Imagery and Explanatory Satisfaction. J. Mind and Behav. 16, 235–253 (1995)
5. Myridis, N.E., Chamzas, C.: Sampling on Concentric Circles. IEEE Trans. Med. Imag. 17(2), 294–299 (1998)

6. Myridis, N.E.: The Image Whole and the Image Eye Diagram [IED]. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 3. AISC, vol. 102, pp. 123–129. Springer, Heidelberg (2011)
7. Sinclair, N., Pimm, D., Higginson, W. (eds.): Mathematics and the Aesthetic. Springer, N.York (2006)
8. Terzidis, K.: Algorithmic Architecture. Architectural Press (2006)
9. Triantaphyllou, E., Felici, G. (eds.): Data Mining & Knowledge Discovery. Springer, Berlin (2006)

# On the Bounds on Optimal Bayes Error in the Task of Multiple Data Sources

Robert Burduk

Department of Systems and Computer Networks,
Wroclaw University of Technology, Wybrzeze Wyspianskiego 27,
50-370 Wroclaw, Poland
robert.burduk@pwr.wroc.pl

**Summary.** The paper considers the problem of pattern recognition when we have multiple data sources. We assume that for each data source there are estimated parameters of statistical distributions. The model of classification is primarily based on the Bayes rule and secondarily on the notion of interval-valued fuzzy sets. The set of possible class-conditional probability density functions is represented by an interval-valued fuzzy set. We consider the case where the uncertainty concerns the mean of Gaussian pdf. In the paper the bound on the optimal Bayess error is presented for a full probabilistic information.

## 1 Introduction

Statistical pattern recognition techniques assume that both the probability distribution of classes and the $d$–dimensional feature vector are represented by known numerical values and that knowledge of data is precise. In practice, we are often faced with two types of uncertainty - randomness and fuzziness. Randomness is described and investigated using methods of the probability theory, and it satisfies and obeys statistical laws. In this type of uncertainty subjective influences are not taken into account. Fuzziness is characterised by non-statistical properties and subjective influences and is dealt with on the basis of fuzzy set theory. There are many cases when the available information is a mixture of randomness and fuzziness. In [1] the pattern recognition problem with fuzzy classes and fuzzy information is formulated and the following three situations are considered:

- fuzzy classes and exact information,
- exact classes and fuzzy information,
- fuzzy classes and fuzzy information.

In this paper we consider the second situation. Additionaly, in our study the fuzzy information describes the uncertainty in class-conditional probability density functions which is represented by interval-valued fuzzy sets.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 201–208.
springerlink.com 

The fuzzy sets theory was introduced in 1965 [2]. Since that time, many new approaches and theories treating imprecision and uncertainty have been proposed [3], [4], [5]. The interval-valued fuzzy sets (IVFSs) were proposed as a natural extension of fuzzy sets [6]. There are fuzzy sets in which the membership degree of each element of the fuzzy set is given by a closed subinterval of the interval $[0,1]$. The usefulness of IVF is discussed in many technical aspects. For example application of IVFSs in pattern recognition is presented in [7], [8], [9].

The classification error is one of the useful measure of the performance of a classifier. Competing classifiers can also be evaluated based on their error probabilities. The probability of error is discussed for a single-stage classifier [10], [11], for a hierarchical classifier [12], [13] and for combining classifiers [14], [15]. This problem is also presented in the context of fuzzy information [1], [16], [17], [18].

In this paper, our aim is to present the error probability for the Bayes classifier. In this model we use exact classes and the interval-valued fuzzy information. We assume that the class-conditional probability density functions is represented by IVFS. The main purpose of this paper is to introduce the theoretic aspect of probability error for the selected model of pattern recognition.

The paper is organized as follows. Section 2 introduces the necessary background and describes the Bayes classifier. In section 3 the introduction to IVFS is presented. In section 4 we present the discusion on probability of error for the selected model of pattern recognition.

## 2 Bayes Classifier

A pattern is represented by a set of $d$ features, or attributes, viewed as a $d$-dimensional feature vector $x \in \Re^d$.

Let us consider a pattern recognition problem, in which the class label $\omega$ is a random variable taking values in the set of class labels $\Omega = \{\omega_1, ..., \omega_c\}$. The *priori probabilities*, $P(\omega_i), i = 1, ..., c$ constitute the probability mass function of the variable $\omega$, $\sum_{i=1}^{c} P(\omega_i) = 1$. Assuming that the objects from class $\omega_i$ are distributed in $x \in \Re^d$ according to the *class-conditional probability density function* $p(x|\omega_i)$, $p(x|\omega_i) \geq 0$, $\forall x \in \Re^d$, and $\int_{\Re^d} p(x|\omega_i)dx = 1, i = 1, ..., c$.

Given the prior probabilities and the *class-conditional probability density functions* we can calculate the *posterior probability* that the true class label of the measured $x$ is $\omega_i$ using the Bayes formula

$$P(\omega_i|x) = \frac{P(\omega_i)p(x|\omega_i)}{p(x)} \tag{1}$$

where $p(x) = \sum_{i=1}^{c} P(\omega_i)p(x|\omega_i)$ is the unconditional likelihood of $x \in \Re^d$.

The equation (1) gives the probability mass function of the class label variable $\omega$ for the observed $x$. The decision for that particular $x$ should be made with respect to the posterior probability.

The "optimal" Bayes decision rule for minimizing the risk (expected value of the loss function) can be stated as follows: Assign the input pattern $x$ to class $\omega_i$ for which the conditional risk

$$R^*(\omega_i|x) = \sum_{j=1}^{c} L(\omega_i, \omega_j) P(\omega_j|x) \tag{2}$$

is minimum, where $L(\omega_i, \omega_j)$ is the loss incurred in deciding $\omega_i$ when the true class is $\omega_j$. The Bayes risk, denoted $R^*$, is the best performance that can be achieved. In the case of the zero-one loss function

$$L(\omega_i, \omega_j) = \begin{cases} 0, \, i = j \\ 1, \, i \neq j \end{cases},$$

the conditional risk becomes the conditional probability of misclassification and the optimal Bayes decision rule is as follows:

$$R^*(\omega_i|x) = \arg\max_i P(\omega_i|x). \tag{3}$$

Let $\Psi^*$ be a classifier that always assigns the class label with the largest posterior probability. The classifier based on the Bayes rule is the following:

$$\Psi^*(x) = \omega_i \quad \text{if} \quad \omega_i = \arg\max_i P(\omega_i) p(x|\omega_i). \tag{4}$$

because the unconditional likelihood $p(x) = \sum_{i=1}^{c} P(\omega_i) p(x|\omega_i)$ is even for every class $\omega_i$.

### 2.1 Bayes Error

The error of $\Psi^*$ is the smallest possible error, called the Bayes error [19]:

$$Pe(\Psi^*) = \int_{\Re^d} [1 - P(\omega_i^*|x)] p(x) dx. \tag{5}$$

It is convenient to split the integral into $c$ integrals, one on each classification region. For this case class $\omega_i^*$ will be specified by the region's label. Then

$$Pe(\Psi^*) = \sum_{i=1}^{c} \int_{\Re_i^*} [1 - P(\omega_i|x)] p(x) dx \tag{6}$$

where $\Re_i^*$ is the classification region for class $\omega_i$, $\Re_i^* \cap \Re_j^* = 0$ for any $i \neq j$ and $\bigcup_{i=1}^{c} \Re_i^* = \Re^d$. Substituting (1) into (6) we have [19]:

$$Pe(\Psi^*) = 1 - \sum_{i=1}^{c} \int_{\Re_i^*} P(\omega_i) p(x|\omega_i) dx. \tag{7}$$

It is the Bayes "optimal" error.

## 3 Interval-Valued Fuzzy Sets

An interval-valued fuzzy set $\bar{A}$ on a universe $X$ is defined as [6]:

$$\bar{A} = \{\langle x, M_{\bar{A}}(x)\rangle : x \in X\} \tag{8}$$

where the function $M_{\bar{A}}(x) : X \to D[0,1]$ defines the degree of membership of an element $x \in X$ to $\bar{A}$. The $D[0,1]$ denote the set of all closed subintervals of the interval $[0,1]$.

A more practical definitions for IVFS can be given as follows:

$$\bar{A} = \{(x, \overline{\mu}(x), \underline{\mu}(x))| \quad \forall x \in X\}, \tag{9}$$

where $\underline{\mu}(x) \leq \mu(x) \leq \overline{\mu}(x), \quad \mu \in [0,1]$.

The upper $\overline{\mu}(x)$ and lower $\underline{\mu}(x)$ membership functions defined the footprint of uncertainty (FOU). The FOU is bounded from above by $\overline{\mu}(x)$ and from below by $\underline{\mu}(x)$. The $\overline{\mu}(x)$ and $\underline{\mu}(x)$ are fuzzy sets, which implies that we can to use fuzzy set mathematics to characterize and work with IVFSs.

## 4 Bayes Classifier with Interval-Valued Class-Conditional Probability Density Functions

In section [2] we present the problem of statistical pattern recognition. We assumed randomness in feature and hypothesis space. Now we present a model of pattern recognition with interval-valued fuzzy information.

In the recognition task we have often to deal with multiple data sources. In this case, we can estimate the unknown parameters of the distribution independently for each data source. Probably, the received probability distributions are not identical. If we assume that parameters of the distribution vary within an interval, we can be express the uncertainty of the values in the context of IVFSs. Fig. 1 shows Gaussian pdfs with uncertain mean $\mu = [\underline{\mu}, \overline{\mu}]$. The set of possible pdfs function is IVFS in which the upper membership function is presented as a thick solid line and the lower membership function is presented as a dashed line (see fig. 1). One of the possible pdf is presented as the solid line. The shaded region is the FOU and can be intepreted as the union of all possible pdfs. In other words this region presents the footprint of uncertainty in estimation of pdf from different sources. The value of possible pdf is an interval bounded by lower and upper membership functions.

In our study we consider the Gaussian pdf with uncertain mean. For this case the upper boundary of the FOU is [20]

$$\overline{h}(x) = \begin{cases} N(x; \underline{\mu}, \sigma), & x < \underline{\mu} \\ N(\underline{\mu}; \underline{\mu}, \sigma), & \underline{\mu} \leq x \leq \overline{\mu} \\ N(x; \overline{\mu}, \sigma), & x > \overline{\mu} \end{cases} \tag{10}$$

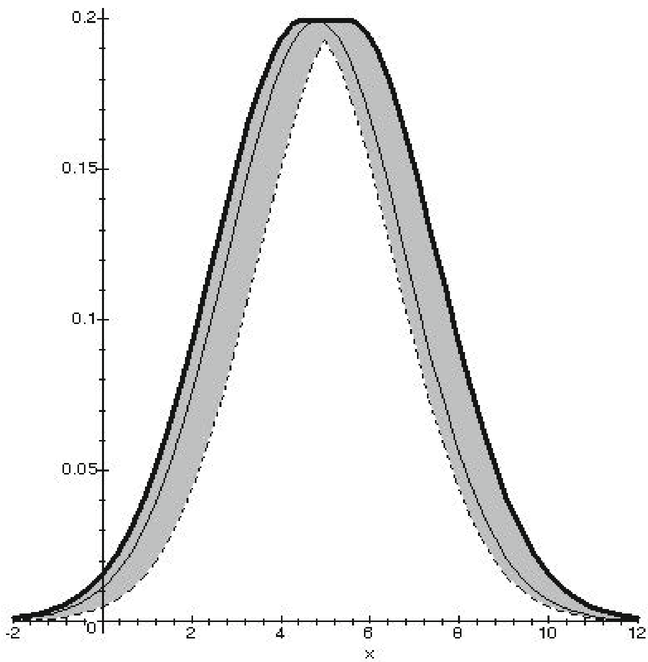



**Fig. 1** The Gaussian pdf with uncertain mean. The shaded region is the FOU.

and the lower boundary of the FOU is

$$\underline{h}(x) = \begin{cases} N(x; \overline{\mu}, \sigma), & x \leq \frac{\underline{\mu}+\overline{\mu}}{2} \\ N(x; \underline{\mu}, \sigma), & x > \frac{\underline{\mu}+\overline{\mu}}{2} \end{cases} \tag{11}$$

Because a one-dimensional Gaussian pdf has 99.7% of its probability mass in the range of $[\mu - 3\sigma, \mu + 3\sigma]$ we have the constrain for the parameter $k \in [0, 3]$. This parameter controls the intervals in which the parameter $\mu$ vary $\underline{\mu} = \mu - k\sigma$, $\overline{\mu} = \mu + k\sigma$. Additionally, parameter $k$ controls the area of the FOU. The bigger the $k$ (the larger the FOU) implies the greater interval $[\underline{\mu}, \overline{\mu}]$ and greater uncertainty of $\mu$.

## 4.1 The Bounds on the Error for the Two Class Problem

Let us consider the binary classifier with a priori probabilities $P(\omega_1) = p$ and $P(\omega_2) = 1 - p$. The class-conditional probability density functions (cpdf) are normal distributions in $\mathbb{R}^1$. For every class we have different source of objects.

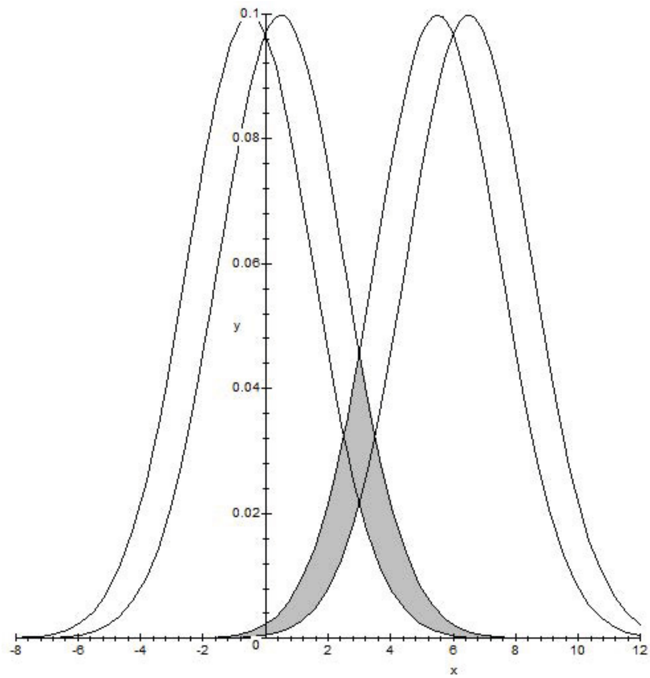



**Fig. 2** The graphical interpretation $\overline{Pe}(\Psi^*) - \underline{Pe}(\Psi^*)$

For this reason we consider that the cpdf are reprecented by IVFS. We assume that the uncertainty concerns the mean of cpdf. For the first class (source) $\mu_1 \in [\underline{\mu}_1, \overline{\mu}_1]$ and for the second class $\mu_2 \in [\underline{\mu}_2, \overline{\mu}_2]$. The standard deviations are the same $\sigma_1 = \sigma_2$. For this assumption we can present the bounds on the error as follows: $\underline{Pe}(\Psi^*) \leq Pe(\Psi^*) \leq \overline{Pe}(\Psi^*)$ where

$$\underline{Pe}(\Psi^*) = \begin{cases} N(x; \overline{\mu}_2, \sigma), & x \leq \frac{\underline{\mu}_1 + \overline{\mu}_2}{2} \\ N(x; \underline{\mu}_1, \sigma), & x > \frac{\underline{\mu}_1 + \overline{\mu}_2}{2} \end{cases}$$

$$\overline{Pe}(\Psi^*) = \begin{cases} N(x; \underline{\mu}_2, \sigma), & x \leq \frac{\underline{\mu}_2 + \overline{\mu}_1}{2} \\ N(x; \overline{\mu}_1, \sigma), & x > \frac{\underline{\mu}_2 + \overline{\mu}_1}{2}. \end{cases}$$

The graphical representation of $\overline{Pe}(\Psi^*) - \underline{Pe}(\Psi^*)$ for the $\mu_1 \in [-0.5, 0.5]$, $\mu_2 \in [5.5, 6.5]$, $\sigma_1 = \sigma_2 = 2$ and $P(\omega_1) = P(\omega_2) = 0.5$ is presented at Fig. 2. The values of $\underline{Pe}(\Psi^*)$ in the function of $|\underline{\mu}_1 - \overline{\mu}_1| = k$ for $p = 0.5, 0.4, 0.3, 0.2, 0.1$ are presented at Fig. 3. These result are recived for $\sigma_1 = \sigma_2 = 2$.

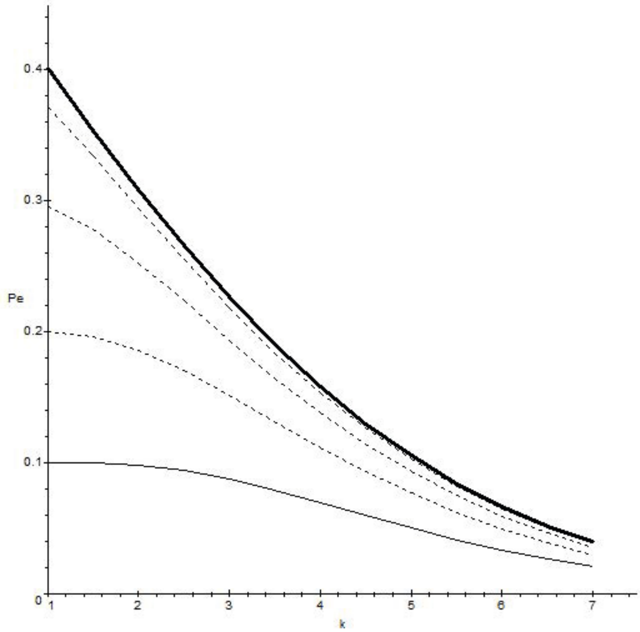



**Fig. 3** The values of $\underline{Pe}(\Psi^*)$ in function of $|\underline{\mu}_1 - \overline{\mu}_1| = k$. Thick solid line - $p = 0.5$, solid line - $p = 0.1$

## 5 Conclusion

In the present paper we have concentrated on the optimal Bayes error. The model of classification is primarily based on the Bayes rule and secondarily on the notion of interval-valued fuzzy sets. These sets represent the uncertainty in class-conditional probability density functions. For this assumptions we presented the upper and lower bound on the optimal Bayes error. The obtained results may be useful in case of different data sources. In future work, the case when the unknown is also the standard deviation should be considered.

**Acknowledgement.** This work is supported in part by the National Science Centre under the grant which is being realized in years 2010–2013.

## References

1. Okuda, T., Tanaka, H., Asai, K.: A formulation of fuzzy decision problems with fuzzy information using probability measures of fuzzy events. Information and Control 38, 135–147 (1978)
2. Zadeh, L.A.: Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications 23, 421–427 (1968)

3. Goguen, J.: L-fuzzy sets. Journal of Mathematical Analysis and Applications 18(1), 145–174 (1967)
4. Pawlak, Z.: Rough sets and fuzzy sets. Fuzzy Sets and Systems 17, 99–102 (1985)
5. Atanassov, K.: Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20, 87–96 (1986)
6. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning - I. Information Science 8, 199–249 (1975)
7. Mitchell, H.B.: Pattern recognition using type-II fuzzy sets. Information Science 170, 409–418 (2005)
8. Zeng, J., Liu, Y.-Q.: Type-2 fuzzy markov random fields and their application to handwritten chinese character recognition. IEEE Transactions on Fuzzy Systems 16(3), 747–760 (2008)
9. Jia, Z., Lei, X., Zhi-Qiang, L.: Type-2 fuzzy Gaussian mixture models. Pattern Recognition 41, 3636–3643 (2008)
10. Antos, A., Devroye, L., Gyorfi, L.: Lower bounds for Bayes error estimation. IEEE Trans. Pattern Analysis and Machine Intelligence 21, 643–645 (1999)
11. Avi-Itzhak, H., Diep, T.: Arbitrarily tight upper and lower bounds on the bayesian probability of error. IEEE Trans. Pattern Analysis and Machine Intelligence 18, 89–91 (1996)
12. Kulkarni, A.: On the mean accuracy of hierarchical classifiers. IEEE Transactions on Computers 27, 771–776 (1978)
13. Kurzyński, M.: On the multistage Bayes classifier. Pattern Recognition 21, 355–365 (1988)
14. Kittler, J.: Combining classifiers: A theoretical framework. Pattern Analysis and Applications 1, 18–27 (1998)
15. Woźniak, M.: Experiments on linear combiners. In: Pietka, E., Kawa, J. (eds.) Information Tech. in Biomedicine. ASC, vol. 47, pp. 445–452. Springer, Heidelberg (2008)
16. Burduk, R.: Classification error in Bayes multistage recognition task with fuzzy observations. Pattern Analysis and Applications 13(1), 85–91 (2010)
17. Pardo, L., Menendez, M.L.: Some bounds on probability of error in fuzzy discrimination problems. European Journal of Operational Research 53, 362–370 (1991)
18. Pardo, J.A., Taneja, I.J.: On the probability of error in fuzzy discrimination Problems. Kybernetes 21(6), 43–52 (1992)
19. Kuncheva, L.I.: Combining pattern classifier: Methods and Algorithms. John Wiley, New York (2004)
20. Zeng, J., Xie, L., Liu, Y.-Q.: Type-2 fuzzy Gaussian mixture models. Pattern Recognition 41, 3636–3643 (2008)

# New Method for Finding Rules in Incomplete Information Systems Controlled by Reducts in Flat Feet Treatment

Jolanta Pauk and Agnieszka Dardzinska

Bialystok University of Technology, Wiejska 45C, 15-351, Białystok, Poland
{jpauk,a.dardzinska}@pb.edu.pl

**Summary.** Flat feet is very popular dysfunction seen in schoolchildren. The effect of different methods of flat feet treatment is not clear. The purpose of this study was using a new algorithm, based on data mining techniques, to predict the success of flat feet correction. The results show that the rules extracted in database are correlated with previous made statistical analysis and doctors suggestions in 82%. Clinicians should reduce the arch height by using physical therapy exercises. Our results show that the arch height correction is increased by age and place of living, and decreased as body mass increased.

## 1 Introduction

Flat feet is very popular dysfunction seen in schoolchildren. The structure of the flat foot is different than in typical subjects because of lower arch height in flat feet subjects. Not all types of flat foot needs a treatment. Flexible flat foot subjects attending the prosthetics and orthotics units come with prescriptions from orthopedic surgeons for arch supports [1]. It also can be treated with orthotic insoles, corrective exercises, and physical activity [2, 3]. However the effect of different methods of flat feet treatment is not clear yet. Wenger [4] has suggested that the flexible flat foot in young children should be self-correcting, and no requiring treatment. In contrast Rose at al. and Pauk et al. [5, 6] suggest that those subjects who fall outside the normal range of parameters require some form of treatment. Nowadays, flat foot treatment methods has provided relief for many patients as well as generated a high volume of medical data in the format of matrix-in-matrix, which is not suitable for traditional statistical or data mining algorithms. On the other hand, due to the fact that objective methods are lacking to detect some of flat foot symptoms, it is also interesting for the clinical doctors to be able to learn the essences of flat foot through the evaluation data. Thus, the authors started their initial research by exploring the relationships among the complex factors of the treatment and recovery patterns in different categories of patients for the purpose of

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 209–214.
springerlink.com 

optimizing the treatment process as well as learning the essence of the children flat foot problems. This study aimed to quantify the effect of such factors as: age, Cole index, place of residence, gender, orthotic treatment for flexible flat foot correction and to investigate long term outcomes. We hypothesized that using data mining techniques may predict the success of flat feet correction. The main purpose of this study is to present a new method to extract rules from incomplete information system for selection the best intervention or flat feet correction.

## 2 Methodology

### 2.1 Subject Description

Sixty eligible plano-valgus and 50 age-matched control children were recruited. Both patients and control subjects were randomly selected from a total population of 450 primary school children. The average age was 12.5±2.2 (range, 7 to 15) years. The evaluation was carried out two times: before treatment and two years post-treatment with orthotic insoles. The local ethics committee approved the study. All parents were received full information about the study before giving signed consent. Subjects were screened with a detailed medical history and were not treated for neither any systemic disease nor flat feet in past. Inclusion criteria were: age range 7-15 years, arch height of bilateral feet, knee and hip position, ankle pain. Exclusion criteria were any other disorders different than flat feet. Subject's body weight was measured using a scale with resolution of 100 g, subject's height was measured by stadiometer. For description of children body mass and height, the Cole index was used [7]. For measuring arch height, the children sat in a chair and placed their feet on level ground. Subjects were tested in random order. Arch height was measured with an optoelectronic system (BTS Smart, Italy). Three reflective markers were placed on foot while subject was seated with the subtalar joint in a neutral position. An experienced physiotherapist placed the markers on the navicular tuberosity, medial aspect of calcaneus, and medial aspect of first metatarsal head. The distance between the floor and the line between the markers on calcaneus and first metatarsal were added afterwards. Information about the duration of wearing orthotic insoles was obtained from questionnaire administered by trained interviewer [8]. The same questionnaire was completed by child and parent separately on the same day. In the same questionnaire children and parents were asked about the duration of corrective exercise [weeks], and the duration of wearing orthotic insoles [months] in the last two years.

### 2.2 Proposition of a New Method to Extract Rules from Incomplete Information System

Our goal is to collect all elements in the incomplete equivalence classes generated by all the condition attributes together. The set formed by such elements is a covering of domain $X$. For any object $x \in X$, $ER(x)$ can be obtained from the covering rough sets model. It follows from the condition attribute value that $ER(x)$ is an undistinguished minimal object set. If no more conditions other than the given condition attribute value exist, then the condition attribute value of all the objects in $ER(x)$ are identical [9, 12]. In the incomplete information system, another important condition should be taken into consideration, which is the value of decision attribute. If the decision attribute value can be determined, the unknown value of classification attribute can be estimated in the following way:

- If $x_i, x_j \in ER(x_k)$ and $(x_i, d) = (x_j, d)$ for decision attribute $d$, then all the condition attribute values of $x_i$ and $x_j$ can be transformed to the known value.
- If $x_i, x_j \in ER(x_k)$ and there exist decision attribute $d$ such that $(x_i, d) \neq (x_j, d)$, then we assume $x_i$, $x_j$ can be distinguished if there exists an estimate of the unknown attribute value, so that the classification attribute value is not identical. When an unknown attribute value of some object is estimated, the corresponding value 0 or 1 is displaced by it. Then the object with the new attribute value is compared with other objects in the common $ER(x)$. If the object can be distinguished with others, then it should be removed from $ER(x)$.

The same steps are repeated up to the fixed point. In such way it is possible to obtain the decision rules according to the objects and their decision attribute value. The algorithm is explained on example presented below. The incomplete information system given is in Table 1.

**Table 1** An incomplete information system

| Objects | Attributes | | |
|---|---|---|---|
| | a | b | d |
| x1 | N | H | - |
| x2 | H | L | - |
| x3 | N | H | + |
| x4 | L | L | 0 |
| x5 | L | H | + |
| x6 | L | N | + |
| x7 | L | | 0 |
| x8 | | H | - |
| x9 | | N | - |

Attributes $a$ and $b$ are classification ones, and attribute $d$ is $a$ decision attribute. Classification attributes can be numerical or symbolical, decision attribute classify objects into one from three classes. The algorithm consists of five main steps as follow:

Step 1 : Denote incomplete equivalence classes of all the

$X/\{a\} = \{\{(x2,1),(x8,0),(x9,0)\},\{(x4,1),(x5,1),(x6,1),$
$(x7,1),(x8,0),(x9,0)\},\{(x1,1),(x3,1),(x8,0),(x9,0)\}$

$X/\{b\} = \{\{(x1,1),(x3,1),(x5,1),(x7,0),(x8,1)\},\{(x2,1),$
$(x4,1),(x7,0)\},\{(x6,1),(x7,0),(x9,1)\}$

Step 2 : Put all elements of each classification attribute of incomplete equivalence classes together.

$\{\{(x2,1),(x8,0),(x9,0)\},\{(x4,1),(x5,1),(x6,1),(x7,1),$
$(x8,0),(x9,0)\},\{(x1,1),(x3,1),(x8,0),(x9,0)\},\{(x1,1),$
$(x3,1),(x5,1),(x7,0),(x8,1)\},\{(x2,1),(x4,1),(x7,0)\},$
$\{(x6,1),(x7,0),(x9,1)\}\}$

The $R(xi)$ is as follows:

$ER(x1) = \{x1,x3,(x8,0)\}$ $ER(x2) = \{x2\}$ $ER(x3) = \{x1,x3,(x8,0)\}$
$ER(x4) = \{x4,(x7,0)\}$ $ER(x5) = \{x5,(x7,0),(x8,0)\}$
$ER(x6) = \{x6,(x7,0),(x9,0)\}$ $ER(x7) = \{(x7,0)\}$ $ER(x8) = \{(x8,0)\}$
$ER(x9) = \{(x9,0)\}$

Step 3 Simplify sets $ER(xi)$. Note that $ER(x)$ is the minimal set of undistinguishable object according to the condition attribute value. In general, if $ER(xj) \subseteq ER(xi)$, then we delete $ER(xj)$. For the example, the simplified $ER(xi)$ is as follows:

$ER(x1) = \{x1,x3,(x8,0)\}$ $ER(x2) = \{x2\}$ $ER(x4) = \{x4,(x7,0)\}$
$ER(x5) = \{x5,(x7,0),(x8,0)\}$ $ER(x6) = \{x6,(x7,0),(x9,0)\}$

Step 4 Evaluate the possible value of the unknown attribute according to the corresponding extracted rule. Since the decision attribute value of the object $x4$ is consistent with that of the object $x7$, we can evaluate that value of attribute $b$ can be $L$. Note that the default attribute value of the object $x7$ has been evaluated, and $ER(x6) = \{x6,(x7,0),(x9,0)\}$ it follows $(x7,0)$ in $ER(x5) = \{x5,(x7,0),(x8,0)\}$ is distinct from the other objects in $ER(x)$, then $(x7,0)$ can be deleted. Similarly, in $ER(x3) = \{x1,x3,(x8,0)\}$ objects $x1$ and $x8$ have the same attribute values, thus we can evaluate possible value of attribute $a$ in $x8$ as $N$. In such case the value $(x8,0)$ in $ER(x5) = \{x5,(x8,0)\}$ is deleted. In $ER(x6) = \{x6,(x9,0)\}$ objects $x6$ and $x9$ have different decision attribute, so the value of attribute a for object $x9$ is still unknown. It means that for object $x9$ we have two values of attribute a with equal weights 1/3: $(aL,1/3),(aN,1/3),(aH,1/3)$ (Table 2).

step 5 : Extract and simplify the rules. For extracting rules from such information system we use algorithm ERID [11], which is similar to LERS [5]. For the above example, the certain rules can be extracted as follows:
$(a,N)*(b,N) \rightarrow (d,-)$ $(a,H)*(b,L) \rightarrow (d,-)$ $(a,H)*(b,N) \rightarrow (d,-)$ $(a,L)*(b,H) \rightarrow (d,+)$ $(a,L)*(b,L) \rightarrow (d,0)$ It is possible to obtain all possible rules in a similar way.

**Table 2** More complete information system

| Objects | Attributes | | |
|---|---|---|---|
| | a | B | D |
| X1 | N | H | - |
| X2 | H | L | - |
| X3 | N | H | + |
| X4 | L | L | 0 |
| X5 | L | H | + |
| X6 | L | N | + |
| X7 | L | L | 0 |
| X8 | N | H | - |
| X9 | | N | - |

## 3 Results

The two years post-treatment group demonstrated that the arch height increases in average by 33% ($p < 0.05$) in flat feet. The average of arch height and Cole index were respectively 14.5 (range, 6 to 23) mm and 107.9 (range, 79.95 to 148.08) post corrective treatment. In control group the average arch height was unchanged and remained at range 33 to 39 mm ($p < 0.05$). The average duration of wearing orthotic insoles was 8.3±4.9 months. The differences between child and parent's reports were not statistically significant ($p > 0.05$). Significant correlation coefficients ($r > 0.80$, $p < 0.05$) was observed, and the agreement between child and parent agreement was substantial ($\kappa > 0.7$). The rules extracted in database are in 82% correlated with previous made statistical analysis and doctors suggestions. Clinicians should reduce the arch height by using for example physical therapy exercises. Our model suggests slightly different interpretation meaning that the arch height correction is increased by age and place of living, and decreased as body mass increased.

## 4 Conclusions

This paper put forward a new algorithm to extract rules from incomplete information system based on the reducts of rough sets models. The presented algorithm estimates some unknown condition attribute values and extract rules from the new, more complete information system. The algorithm was implemented and initially tested on flat feet children database. The results are very promising. We proposed new simple model to estimate arch height correction in flat-foot children based on rules extraction.

## References


1. Leung, A.K.L., Mak, A.F.T., Evans, J.H.: Biomechanical gait evaluation of the immediate effect of orthotic treatment for flexible flatfoot. Prosthetics and Orthotics International 22, 25–34 (1998)
2. Daunoravičienė, K., Pauk, J., Raso, J., Griškevičius, J.: Research of Muscular Activity during Gait of Persons with Cerebral Palsy. In: Choraś, R.S., et al. (eds.) Image Processing and Communications Challenges 2. AISC, vol. 84, pp. 255–261. Springer, Heidelberg (2010)
3. Pauk, J., Daunoraviciene, K., et al.: Analysis of the plantar pressure distribution in children with foot deformities. Acta of Bioengineering and Biomechanics 12(1), 29–34 (2010)
4. Wenger, D.R., Diego, S., et al.: Corrective shoes and inserts as treatment for flexible flatfoot in infants and children. J. Bone Joint Surg. 71A, 800–810 (1989)
5. Rose, G.K., Welton, E.A., Marshall, T.: The diagnosis of flat foot in the child. J. Bone Joint Surg. 67B, 71–78 (1985)
6. Pauk, J., Ezerskiy, V.: The effect of foot orthotics on arch height: Prediction of arch height correction in flat foot children. Biocybernetics and Biomedical Engineering 31(1), 51–62 (2011)
7. Cole, T.J., Belizzi, M.C., et al.: Establishing a standard definition for child overweight and obesity worldwide: international survey. Br. Med. J. 320, 1–6 (2000)
8. Kowalski, K.C., Crocker, P.R.E., Faulkner, R.A.: Validation of the Physical Activity Questionnaire for Older Children. Pediatric Exercise Science 9, 74–186 (1997)
9. Ras, Z., Dardzinska, A.: Extracting Rules from Incomplete Decision Systems: System ERID, pp. 143–154. Springer (2005)
10. Chmielewski, M.R., Grzymala-Busse, J.W., et al.: The rule induction system LERS-A version for personal computers. Found. Comput. Decision Sci. 18, 181–212 (1993)
11. Kryszkiewicz, M.: Rough set approach to incomplete information systems. Information Sciences, 39–49 (1998)
12. Dardzińska, A., Raś, Z.W.: $CHASE_2$ – Rule Based Chase Algorithm for Information Systems of Type $\lambda$. In: Tsumoto, S., Yamaguchi, T., Numao, M., Motoda, H. (eds.) AM 2003. LNCS (LNAI), vol. 3430, pp. 255–267. Springer, Heidelberg (2005)
13. Dardzinska, A., Ras, Z.: On Rule Discovery from Incomplete Information Systems. In: Proceedings of ICDM 2003 Workshop on Foundations and New Directions of Data Mining, pp. 31–35 (2003)

# OTO Model of Building of Structural Knowledge – Areas of Usage and Problems

Krzysztof Wójcik

Pedagogical University of Cracow, Institute of Computer Science
ul.Podchorążych 2, 30-084 Kraków, Poland

**Summary.** This article describes an OTO (Observation-Transformation-Operation) model which allows to improve building of the knowledge structure of the simple agent systems. The presented approach tries to overcome the crucial problems of the task of the automatic ontology building. To this end inductive learning methods and knowledge transformations are utilized. The article provides a brief outline of various forms of these transformations. The chosen example of their usage in building of the partial knowledge structure is also presented. As a conclusion, the paper points to the many possible areas of the model usage, mainly in the field of the image processing and image understanding.

## 1 Introduction and Bibliography Remarks

The interpretation and understanding of the images most often require the usage of the knowledge concerning the observed phenomena. Besides advantages, mainly consisting in the system ability to adaptively operate in a changing environment, this approach faces several difficulties. Among them one of the most important is the assumption that a useful model of phenomena is available. Its creation requires a definition of concepts and relationships between them [3]. This process strongly depends on the knowledge of the researcher. In order to overcome these problems the article proposes a usage of the general scheme of building a knowledge structure with the ability to learn, called an OTO (Observation-Transformation-Operation). The widest theoretical background of this approach is provided by [7]. In this article only a very brief outline of it is given. We can notice some connections between the described approach and well-known methods. We may find them in the field of:

1. Image Recognition and Image Understanding Techniques (IU).
We can notice the strong connection between the proposed methodology and the syntactic variants of the IU, especially described in the [6]. The hierarchical

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 215–222.
springerlink.com 

graph methods e.g. [5] relates also to the hierarchical knowledge structure described in the article.

2. Learning from the Observations, Inductive Learning Methods.

As a source of knowledge the presented scheme utilizes the observation of the relationships between objects in the given reality. This approach is widely described in [3].

3. Automatic Ontology Building, Knowledge Engineering Methods.

Many of terms introduced by the methodologies of the Automatic Ontology Building [1] relate to elements of our approach (e.g. the main term -"ontology" relates to the "hierarchical structure of concepts" - section 2). We may also notice the connections between the presented in the article approach and elements of the OWL (Web Ontology Language). It is the knowledge representation language that describes the features of the classes, relationships, and defines the operations on the ontology structures. The analogous operations concerning the knowledge structure are a base of presented method.

## 2 Inductive Scheme of the Knowledge Building

At the beginning, some theoretical background needs to be provided. We assume that the phenomena from the given reality consist of a elementary parts, which we will call primitive objects or instances (set of them will be denoted by $\mathbf{X}_0$). The objects correspond to the primitive "concepts", also called "types" ($\mathbf{C}_0$ - set of them). Additionally, the objects are connected by primitive relationships ($\mathbf{D}_0$ - set of them).

The concepts are the generalization of objects, we presume that the presence of certain objects in the given reality is a condition of the creation of a new concept. Morever, the objects which will be used in concept creating should be in some way significant. We assume that this will happen if the objects are connected by the relationships. Consider one of them, denoted by:

$r_i$, where: $r_i \in \mathbf{D}_0$, $i \in \mathbf{I}$, $\mathbf{I}$ is a set of indices of relations.

By $t_{ik}$ we will denote a $n$-tuple, which satisfies $r_i$ relation,

where: $t_{ik} \in \mathbf{X}_0^n$ (n$th$ Cartesian power of $\mathbf{X}_0$), $i \in \mathbf{I}$, $k \in \mathbf{K}$, $\mathbf{K}$ is a set of indices of tuples.

The selected tuples will be a base of the new concept construction. Let us try to construct the set (group) of them. To each tuple we should attach the index that identifies the relationship which is satisfied by it. The group $\mathbf{G}$ will be defined as an ordered set of pairs:

$\mathbf{G} = \{(t_{ik}, i): \quad i \in \mathbf{I},\ k \in \mathbf{K}\}$ where: $(t_{ik}, i)$ - pair contains the tuple and the index of the relationship.

Let us transform now the $\mathbf{G}$ set by the simple replacing of each object in each tuple with the label of the object type. We obtain a set, denoted by $\mathbf{S}$, that describes an abstract arrangement of relations in the group. The groups that have identical or similar (according to some distance function) arrangements will be regarded as similar. Among the set of the similar group

we may determine one, most characteristic group. Let us denote by $\mathbf{G}_p$ this group and by $\mathbf{S}_p$ its arrangement (pattern) of relations (see the more complete problem description - [7]). Based on the $\mathbf{S}_p$ set the new concept will be defined. In this brief description we omit the additional problems concerning that definition. Let us only note them: problem of the definition of the attributes of the objects of the new type (this process is denoted by **FA**), problem of creating the new relationships between objects (denoted by **FR**). Finally, the concept is defined as a composition of elements: ($\mathbf{S}_p$, **FR**, **FA**).
After a concept creation we can create objects of the new type. The processes of concept and object creating may be operated repetitively. In each iteration the new concepts and objects may be joined to the created earlier ones. (the extended sets of concepts and objects will be denoted by $\mathbf{C}, \mathbf{X}$ respectively).

The whole process leads to the construction of hierarchical structure of concepts and hierarchical structure of objects. The structure of concepts includes the new general knowledge and can be treated as a model of the reality.

The object structure (created on the basis of the input object set) contains knowledge about the particular scene. We may consider the presence of the objects of the certain types in that structure as a system response, which can be employed in the tasks of the scene interpretation, control etc.

Let us notice two important matters. The first is associated with the problem of evaluation of the created knowledge (see [8]). In this short description we point to a few general approaches:
Evaluation by the practical action (the system is treated as an agent acting in a certain environment)
Evaluation by the directly defined performance function.
Evaluation by the usage of the arbiter, teacher (supervised learning).

The second problem is finding a suitable concept structure. There are many possible variants of $\mathbf{FR}, \mathbf{FA}$ transformations, thus the solution is a point of an extremely large search space. This is a central problem of the whole presented method. The paper [8] provides some proposition of its solution. For example: certain kind of evolutionary algorithm, certain kind of techniques called local-search, supported by the heuristic approaches (the bundle search methods).

Let us go back to the process of concept creating. The main steps of our method: calculating all possible relationships between objects, determining the $\mathbf{S}_p$, performing the $\mathbf{FR}, \mathbf{FA}$ transformations and finally the creation of new objects, may be described in a more general form as a sequence:

**A.** The observation.
In this process the relationships between objects are identified, it provides the answer to the question: what is visible at present? i.e. in this step.

**B.** The transformation.
This process refers to the transformation of the knowledge. Its goal is to change of knowledge in order to better describe the observed phenomena in the future (i.e. in the next steps of the process). The transformed knowledge could take into consideration such relationships which are not directly observed.

**C.** The operation.
The essence of this process is the usage of knowledge to change the reality. The operation may lead to the creation or removing of the objects.

The iterative process presented above may be considered as a kind of framework, let it be denoted by an OTO (Observation-Transformation-Operation). By the help of the OTO model we can easily identify the significant variants of the whole method. Let us notice here a few most interesting ones.

**A.** The observation.
In a broader sense of the observation process the system may refer to everything that brings in information. Besides the objects it may be the concept as a general being. The system may observe the relationships, for example the similarity, between two (or more) concepts. This makes it possible to build a new concept after the identification of an analogy in the relationships among the objects.

Computation of all relationships among objects and creating the $\mathbf{G}$ sets may be greatly simplified when we assume that one specific object called the "central object" exists. This object is is an argument of all relations in the $\mathbf{G}$ set. Consider one chosen object, then we may easily identify all relations that are satisfied by it. In the case of considering only binary relationships, the complexity of this process grows with the number of objects like $O(n^2)$ (for each potential central object from $\mathbf{X}$ we calculate relationships to other objects from $\mathbf{X}$).

**B.** The transformation.
The obvious kinds of the knowledge transformation may lead to the replacement of one (or more) relationships in the $\mathbf{S}_p$ with the other one (ones). The other sort of the transformation refers to a change of the number of objects that meet the relationship. Let us consider a binary relationship $r_0$. We may imagine the new unary relation $r_1$ that is satisfied by the particular object $\mathbf{x}$ if the relationship $r_0$ holds between the $\mathbf{x}$ and only one other object from $\mathbf{X}$ set. Similarly, we may define the new relationship $r_2$ which will be satisfied by the $\mathbf{x}$ if $r_0$ is satisfied by the $\mathbf{x}$ and all the objects from $\mathbf{X}$. In the same exact way we may define $r_3$ if $r_0$ is satisfied by the $\mathbf{x}$ and none of the objects from $\mathbf{X}$. This kind of modification will be called "excluding transformation".

Let us notice the other simple transformation of the $\mathbf{S}_p$ set. We may propose a transformation that involves the change of the possible types of the relationship arguments. One situation is especially interesting. Assume that according to some pattern $\mathbf{S}_p$, we are creating a concept which will be denoted by number $m$. Technically speaking it is possible to change certain types of arguments of relations in the $\mathbf{S}_p$ also into $m$. In this way we obtain a "recursive" type which is defined by itself.

The described above transformations concern one particular set $\mathbf{S}_p$. However, from this set we may create a copy, which will be transformed. In this way the two (or more) $\mathbf{S}_p$ sets may create an alternative concept definitions (see the example in the next section).

**C.** The operation.
Let us notice here only two main variants of the object creation: the creation of all possible objects and only a reduced set of the most dissimilar ones.

In the next section we will deal with particular example of the OTO model usage.

## 3 Example of the Knowledge Structure Building

Let us assume that we want to build a knowledge about the operations on integer numbers. We must give the description of the concept "number" before. This concept may be defined on the basis of a more elementary concept of "equinumerosity of sets". We are not interested in the well-known definition of this concept which uses a bijection transformation. We want to define it with minimal assumptions and knowledge about the observed objects.

At the beginning we have to define the initial sets $\mathbf{X}_0, \mathbf{C}_0, \mathbf{D}_0$. As a primitive concept we consider a class of vectors, which have only one essential feature called *color*. This attribute may have two values: 0 and 1 (*blue* and *red*). According to *color* value all the primitive objects can be divided into two sets. There are also defined primitive binary relationship: *equality* and *inequality* of the *color* as well as two unary relations *having_ blue_ color* and *having_ red_ color* (all relations are denoted by the numbers: 13,15,5,6 respectively, see below). In our example we apply the following variants of the OTO subtasks (the meanings of the terms could be found in the previous chapter): single-time objects usage; central object method with the binary relationships; the excluding transformation; the creation of the recursive type; the creation only of a reduced set of the most dissimilar objects.

The knowledge structure is built according to one simple example of object sets containing four *blue* objects and three *red* ones. The evaluation of the created knowledge structure is performed by testing if it gives a proper output (objects of certain type) for ten testing examples.

The searching through all the nodes of the decision tree is supported by the bundle search methods with use the local criterion function. This function, which assesses the ability of the pattern to create the useful knowledge, combines only two elements: the count of the appearance of the **G** set and a "age" of the concepts which create considered **G** set. The feature "age" is determined by the type of central object in the pattern (the higher type number - the later creation of the concept). On the each stage of searching process several knowledge transformation are checked, e.g.: the excluding transformation of all binary relationships in the pattern; the creation of the recursive concept (stop criterion of the object creation - the newly created object includes itself).

The execution of the searching process was performed in simple environment of a test program. It has generated hundreds of solutions which contain proper knowledge structures. We will present one chosen solution only.


**Fig. 1** The part of the screen of the test program, concept stucture decrtiption

The Figure 1 illustrates a part of the screen captured during the run of the test program.

The records in the below window represents the discovered concept structure after three steps of the OTO loop. We interpret that notation as follows:

typ 20 q1 pat:

typ 20 - a header of the definition of concept (type) no. 20, this is the concept of the primitive objects, q1 - a performance value of this concept, which is of no importance in our discussion, pat: - the start of the relation definitions (empty here);

typ 21 q1 pat: 13: 20 20

typ 21 - the header of the definition of concept 21 (headers will be omitted in further descriptions),

13: 20 20 - the relation 13 (*inequation*) held between two objects of type 20, the object of concept 21 is a pair of the primitive objects (type 20) which have a different color, such a pair will be called a "different pair"; in this denotation the first argument of the relation corresponds to a "central object";

typ 22 q1 pat: 15: 21 21

15: 21 21 - the relation 15 (*equality*) held between two objects of type 21, the object of concept 22 contains two objects of type 21, i.e. two different pairs;

typ 22 q1 pat: 15: 22 21

15: 22 21 - the relation 15 (*equality*) held between two objects of type 22 and 21, this is a transformation of the previous definition, type 22 becomes the "recursive type", the object of the extended concept 22 may be a set of the different pair (type 21);

typ 24 q4 pat: neg 13: 22 20 15: 22 20 15: 22 21
this is the excluding transformation (denoted by "neg") of the relations:
13: 22 20 15: 22 20 15: 22 21,
object of type 24 may be interpreted as a set of different pairs for which:
13:22 20 - another object (type 20) having the different color does not exist,
15:22 20 - another object (type 20) having the equal color does not exist,
15:22 21 - another pair does not exist (all the pairs have *blue* color),
the last conditions mean that every object of type 24 includes all the different pairs, and does not exist any other object which is not a component of it, so, the possibilities of the creation of the object of the type 24 means that the two sets have the same cardinality.

The created concept structure quite well matches the human manner of the understanding of the concept "equinumerosity". A lot of people checking if two sets have the same number of components construct the pairs of elements belonging to two sets. We can also notice another similarity. We aware that the large hypothesis space size causes the problems of finding the proper concept structure. These problems also involve the human being. We will present an example provided by Educational psychology [2].

It deals with teaching children (aged about five) of the concepts of the set and set equinumerosity. The children are asked to point to equinumerous sets of several objects showed in a picture. Most of the kids give the correct answer, pointing to the sets of five apples and five elephants. However, there are some who have doubts: the apples are small, the elephants are big animals - they say. They are asked by the teacher to draw lines from elephants to apples (visualization of the pairs). Nevertheless, this changes nothing in the child's thinking. "The elephants are big, the apples are small, but additionally there are several lines"[1]. We can clearly see here the troubles in generating the new concept due to a big number of objects, a big number of possible relationships between them and taking into consideration irrelevant object attributes.

In this brief paper we will not present other, very interesting solutions. Most of them seem to be complicated and/or strange. Despite this, they work correctly.

The OTO approach may be practically used in various tasks of the image processing. Especially in the method of the image quality enhancement. Assume, we want remove the noise effects, artifacts or dust seeds from the picture. The main problem here is how to recognize the little obstructions among other small objects on the picture. The simple image recognition methods consider the properties of the particular objects, but it is not enough to perform a correct classification. The other small objects have often the same properties. It is visible, that we can utilize here the description of the picture provided be the OTO structure of the objects. In that idea the crucial problem is how to evaluate the created knowledge [8].

[1] This is the author's translation of the polish sentence from [2].

## 4 Conclusions

The application of the multi-stage schema Observation-Transformation-Operation in the task of building the knowledge structure seems to be much profitable. Especially the split of the task into distinct stages of creating the particular concepts makes it possible to reach some solutions (the proper knowledge structure). Let us summarize the main advantage of the proposed OTO scheme:

- The possibility of the usage of the heuristic method which utilizes the bundle search and local optimization.
- The opportunity to use the wide combination of knowledge transformation.

Because of some features of the OTO model its utilizing may be very promising. Especially in the areas:

- The relatively "small" autonomic system with the ability to learn. For example systems which are used in image processing task [8].
- The field of humanities. An aid to the human learning process (the automatic search of the concept structure that helps the student understand complex concepts), modeling of the human knowledge structure.
- The building or the support of the building of the ontology systems (expert systems, data bases).

## References

1. Davies, J., Studer, R., Warren, P.: Semantic Web Technologies Trends and Research in Ontology-based Systems. John Wiley & Sons Ltd. (2006)
2. Gruszczyk-Kolczyńska, E., Zielińska, E.: Dziecięca matematyka. Edukacja matematyczna dzieci w domu, w przedszkolu i szkole. WSiP Warszawa (1997) (in Polish)
3. Mitchell, T.M.: Machine Learning. McGraw-Hill Science (1997)
4. Muggleton, S.H., De Raedt, L.: Inductive logic programming: Theory and methods. Journal of Logic Programming 19/20 (1994)
5. Piekarczyk, M., Ogiela, M.R.: Hierarchical Graph-Grammar Model for Secure and Efficient Handwritten Signatures Classification. Journal of Universal Computer Science 17 (2011)
6. Tadeusiewicz, R., Ogiela, M.R.: Medical Image Understanding Technology. STUDFUZZ, vol. 156. Springer, Heidelberg (2004)
7. Wójcik, K.: Inductive Learning Methods in the Simple Image Understanding System. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part I. LNCS, vol. 6374, pp. 97–104. Springer, Heidelberg (2010)
8. Wójcik, K.: Hierarchical Knowledge Structure Applied to Image Analyzing System - Possibilities of Practical Usage. In: Tjoa, A.M., Quirchmayr, G., You, I., Xu, L. (eds.) ARES 2011. LNCS, vol. 6908, pp. 149–163. Springer, Heidelberg (2011)

# Part II

# Communications

# Anomaly Detection Preprocessor for SNORT IDS System

Łukasz Saganowski, Marcin Goncerzewicz, and Tomasz Andrysiak

Institute of Telecommunications, University of Technology & Life Sciences in Bydgoszcz ul. Kaliskiego 7, 85-789 Bydgoszcz, Poland
{lukasz.saganowski,tomasz.andrysiak}@utp.edu.pl

**Summary.** In this paper we propose anomaly detection preprocessor for SNORT IDS Intrusion Detection System [1] base on probabilistic and signal processing algorithms working in parallel. Two different algorithms increasing probability of detecting anomalies in network traffic. 25 network traffic features were used by preprocessor for detecting anomalies. Preprocessor calculated Chi-square statistic test and energy from DWT Discrete Wavelet Transform subband coefficients. Usability of proposed SNORT extension was evaluated in local LAN network.

## 1 Anomaly Detection Algorithms

Intrusion Detection Systems (IDS) are based on mathematical models, algorithms and architectural solutions proposed for correctly detecting inappropriate, incorrect or anomalous activity within networked systems. Intrusion Detection Systems can be classified as belonging to two main groups depending on the detection technique employed: anomaly detection and signature-based detection.

Anomaly detection techniques, which we focus on in our work, rely on the existence of a reliable characterization of what is normal and what is not, in a particular networking scenario. More precisely, anomaly detection techniques base their evaluations on a model of what is normal, and classify as anomalous all the events that fall outside such a model.

### 1.1 Statistical Features

For anomaly detection in network traffic probabilistic techniques can be used [2, 3, 4, 5]. Random variables are created during observations of $p$ variables at time $t$ - $\mathrm{X} = (\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_p)$. Random variables are created based on network traffic features shown in Table 1. The Chi-square multivariate test for Anomaly Detection Systems can be represented by equation 1:

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 225–232.
springerlink.com 

$$X^2 = \sum_{i=1}^{p} \frac{(\mathrm{X}_i - \bar{\mathrm{X}}_i)^2}{\bar{\mathrm{X}}_i}, \tag{1}$$

where $\bar{\mathrm{X}} = (\bar{\mathrm{X}}_1, \bar{\mathrm{X}}_2, ..., \bar{\mathrm{X}}_p)$ is the sample mean vector.

Using only the mean vector in Equation 1 causes that Chi-square multivariate test detects only the mean shift on one or more of the variables. For detecting anomalies in network traffic first we have to create vectors of $\bar{\mathrm{X}}$. Vectors of $\bar{\mathrm{X}}$ are used for creation of "normal traffic" profiles (normal profiles have to be created from network traffic without anomalies). We used Chi-square test because it is not computationally complex. It is an important feature because we have to calculate statistics in real time from network traffic features.

### 1.2 Discrete Wavelet Transform Features

Second algorithm used for anomaly detection is based on Discrete Wavelet Transform[6, 7]. The main goal of wavelet transform is to decompose the input signal into family of some specific functions called wavelets. Wavelets are functions that are generated through a process of dilations and translations of one single function, which is usually named Şmother waveletŤ. The idea of wavelet transform was defined by J. Morlet [8]

$$W_d f(m,n) = \sum_{m,n} f(x) \cdot \Psi_{m,n}(x) \tag{2}$$

where $\Psi_{m,n}$ means a family of discrete wavelet functions. The discrete wavelet transform is computed by applying a separable $1-D$ filter bank to the input signal (see Figure 1). Given a signal $s$ of length $N$, the DWT consists of $log_2 N$ stages at most. The first step produces, starting from $s$ two sets of coefficients: approximation coefficients $cA_1$, and detail coefficients $cD_1$. These vectors are obtained by convolving $s$ with the low-pass filter $Lo\_D$ for approximation, and with the high-pass filter $Hi\_D$ for detail, followed by dyadic decimation (downsampling). The next step splits the approximation coefficients $cA_1$ into two parts using the same scheme, replacing $s$ with $cA_1$ and producing $cA_2$ and $cD_2$ and so on. Wavelet transformation with 3 decomposition levels is presented in Figure 2.

The wavelet decomposition of the signal $s$ analyzed at level $j$ has the following structure: $[cA_j, cD_j, ..., cD_1]$.

In case of proposed ADS system signal represents parameters of network traffic (see Table 1). For detecting anomalies in ADS we are using as a parameter energy of DWT coefficients:

$$Ensub_i = \sum_{n=1}^{K} cP_i^2(n) \tag{3}$$

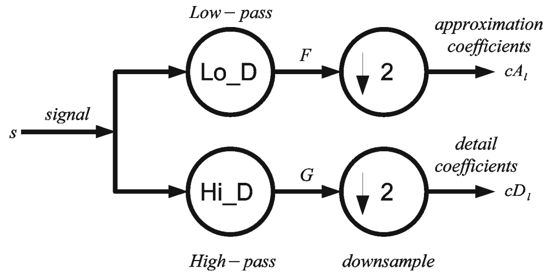



**Fig. 1** Practical realisation of 1-D Discrete Wavelet Transform

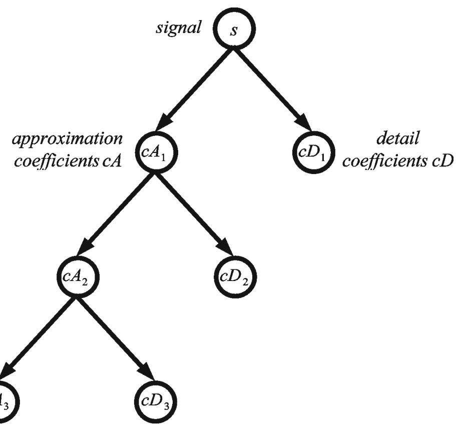



**Fig. 2** Wavelet transformation with 3 decomposition levels

where $P_i$ - DWT coefficients of approximation or detail subbands. Profiles are built from approximation $cA$ and detail $cD$ coefficients calculated during 3 level DWT decomposition.

Additionally, we added to preprocessor calculation of wavelet transform by using Lifting Scheme [9]. Lifting Scheme is an efficient implementation of Wavelet decomposition, where the number of operation can be reduced by a factor of two.

## 2 Proposed ADS Preprocessor for SNORT IDS

In Figure 3 block diagram of proposed ADS SNORT preprocessor is presented. ADS preprocessor (written in C language) is an extension to SNORT IDS system. We are using SNORT as a kind of sniffer which provides different traffic features (see Table 1). For detecting anomalies we are using Chi-square statistic test and coefficients calculated from Discrete Wavelet Decomposition.

For both algorithms first we have to built normal traffic profiles (see stepped lines in Figures 4-5). Traffic profiles were calculated from network traffic collected in 6 weeks. So far the preprocessor was tested with the use of small LAN network to prove usefulness of proposed ADS algorithms (in the next step we are planning to connect our preprocessor to large university network where we will be able to test traffic redirected to our preprocessor from many LAN networks). Network traffic was analyzed in 10 minutest windows [10] for each 25 features. Analysis window can be arbitrarily set during preprocessor start.

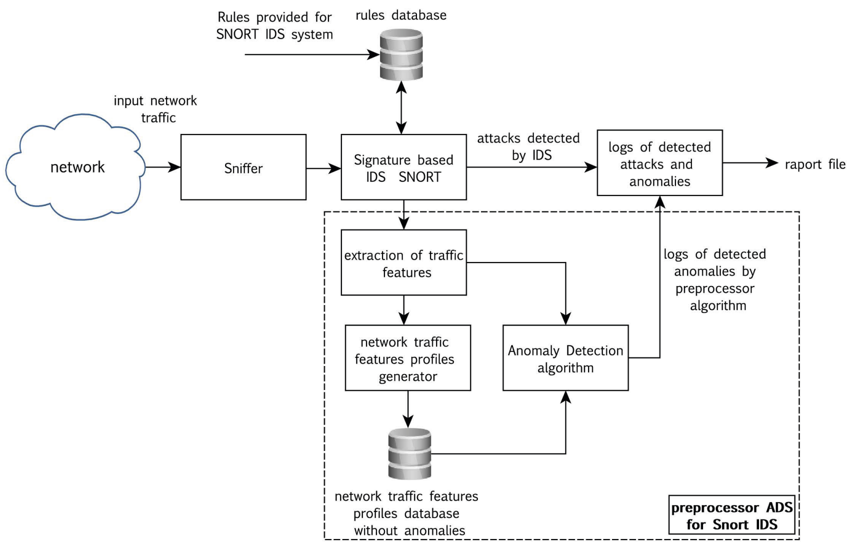


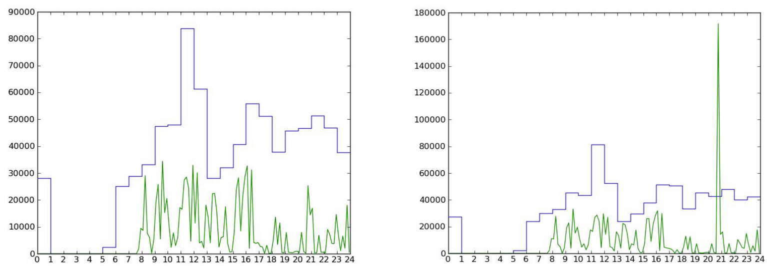

**Fig. 3** Anomaly Detection preprocessor block diagram for SNORT IDS system

Traffic profiles are stored in local preprocessor database (preprocessor can update traffic profiles in any time by switching preprocessor to appropriate mode). During normal work preprocessor calculates Chi-square test and energies from DWT subbands coefficients and compare it to the network profiles collected in database. Preprocessor indicates anomalies when the parameters calculated by ADS algorithms during normal work exceed boundaries designated by interval $< \mu - 3\sigma, \mu + 3\sigma >$
where: $\mu$ - is a mean calculated for one analysis window for a given network profile; $\sigma$ - standard deviation calculated for one analysis window for a given network profile.

When the preprocessor indicates anomaly a report is generated to a log file. A log file consist of the information when anomaly starts and ends, the traffic feature which indicates alarm and about confidence level of a given alarm.

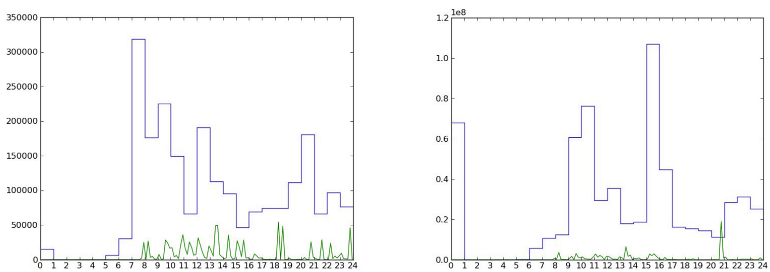
**Table 1** Network traffic features captured by SNORT ADS preprocessor

| | | | |
|---|---|---|---|
| $f_1$ | 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 | | |




(a) (b)

**Fig. 4** Chi-square profile (stepped line) and value of Chi-square 24 hour test (axis $x$ - time in hours) for current network traffic (a) without anomaly (b) with anomaly




(a) (b)

**Fig. 5** DWT coefficient energy profile (stepped line) and value of DWT energy coefficients during 24 hour test (axis $x$ - time in hours) for current network traffic (a) without anomaly (b) with anomaly

Exemplary profiles (stepped line) together with real time calculated values (values of Chi-square test and energy from DWT coefficients) during normal preprocessor work are shown in Figures 4-5.

## 3 Experimental Results

Proposed ADS preprocessor was evaluated with the use of local LAN network. The preprocessor examined summary traffic from entire subnet. Usability of proposed solution was evaluated by simulating different attack on tested LAN network. We used Back Track [11] Linux distribution for simulating different attack such as eg. various port scanning, DoS, DDoS, Syn Flooding, pocket

**Table 2** Detection Rate DR [%] achieved by subsequent network traffic features

| Feature | Chi-sqaure | Mallat | Lifting scheme | Feature | Chi-sqaure | Mallat | Lifting scheme |
|---|---|---|---|---|---|---|---|
| $f_1$ | 5.26 | 5.26 | 5.26 | $f_{14}$ | 0.00 | 5.26 | 10.52 |
| $f_2$ | 5.26 | 10.52 | 10.52 | $f_{15}$ | 0.00 | 10.52 | 10.52 |
| $f_3$ | 0.00 | 10.52 | 10.52 | $f_{16}$ | 0.00 | 0.00 | 0.00 |
| $f_4$ | 15.78 | 10.52 | 10.52 | $f_{17}$ | 5.26 | 5.26 | 5.26 |
| $f_5$ | 10.52 | 10.52 | 10.52 | $f_{18}$ | 10.52 | 10.52 | 10.52 |
| $f_6$ | 0.00 | 0.00 | 0.00 | $f_{19}$ | 5.26 | 5.26 | 10.52 |
| $f_7$ | 0.00 | 0.00 | 0.00 | $f_{20}$ | 10.52 | 5.26 | 5.26 |
| $f_8$ | 15.78 | 31.58 | 31.57 | $f_{21}$ | 5.26 | 10.52 | 10.52 |
| $f_9$ | 94.73 | 94.73 | 84.21 | $f_{22}$ | 0.00 | 0.00 | 0.00 |
| $f_{10}$ | 73.68 | 94.73 | 78.95 | $f_{23}$ | 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 | 15.78 | $f_{25}$ | 5.26 | 0.00 | 0.00 |
| $f_{13}$ | 10.52 | 10.52 | 10.52 | | | | |

**Table 3** False Positive FP [%] achieved by subsequent network traffic features

| Feature | Chi-sqaure | Mallat | Lifting scheme | Feature | Chi-sqaure | Mallat | Lifting scheme |
|---|---|---|---|---|---|---|---|
| $f_1$ | 4.46 | 7.43 | 8.96 | $f_{14}$ | 3.73 | 7.48 | 9.64 |
| $f_2$ | 4.07 | 7.99 | 9.42 | $f_{15}$ | 3.91 | 7.17 | 9.32 |
| $f_3$ | 4.49 | 7.96 | 9.69 | $f_{16}$ | 0.02 | 0.02 | 0.02 |
| $f_4$ | 4.24 | 6.06 | 6.90 | $f_{17}$ | 0.34 | 0.39 | 0.39 |
| $f_5$ | 4.57 | 5.62 | 3.94 | $f_{18}$ | 3.90 | 8.74 | 9.95 |
| $f_6$ | 2.86 | 4.14 | 2.23 | $f_{19}$ | 4.37 | 8.36 | 10.26 |
| $f_7$ | 5.18 | 5.33 | 5.98 | $f_{20}$ | 3.71 | 8.50 | 9.95 |
| $f_8$ | 4.20 | 8.28 | 8.62 | $f_{21}$ | 3.81 | 7.09 | 9.11 |
| $f_9$ | 6.69 | 9.13 | 0.05 | $f_{22}$ | 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 | 12.64 | $f_{24}$ | 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 | 8.14 | | | | |

fragmentation and others. When the preprocessor indicates a possible anomaly a log is generated to a text file. Granularity of analysis depends on window analysis time. Time of analysis window can be arbitrarily set during start of preprocessor.

Detection rate $DR$ and false positive $FP$ for 25 traffic features were presented in Table 2 and Table 3. The best results were achieved for features $f_9$ and $f_{10}$. Detection rate and false positive depends on ADS algorithm and calculated traffic feature. $DR[\%]$ for $f_9$ and $f_{10}$ changes in boundaries $73.68 - 94.73$ in turn $FP[\%]$ has values between $0.05 - 9.13$.

## 4 Conclusion

This paper presents proposition of anomaly detection preprocessor for SNORT IDS system. The major contributions are: proposition of new SNORT preprocessor where at the same time two different algorithms (Chi-square and DWT) were used to detect anomalies. The preprocessor was examined in real network. The presented results prove that the presented algorithms can be used for improving cybersecurity and resilience of the network infrastructures.

## References


1. SNORT IDS, http://www.snort.org/
2. Ye, N., Chen, Q., Emran, S.M.: Chi-squared statistical profiling for anomaly detection. In: Proc. IEEE SMC Inform. Assurance Security Workshop, West Point, pp. 182–188 (2000)
3. Scherrer, A., Larrieu, N., Owezarski, P., Borgant, P., Abry, P.: Non-Gaussian and Long Memory Statistical Characterizations for Internet Traffic with Anomalies. IEEE Trans. on Dependable and Secure Computing 4(1) (2007)
4. Choraś, M., Saganowski, Ł., Renk, R., Hołubowicz, W.: Statistical and signal-based network traffic recognition for anomaly detection. Expert Systems: The Journal of Knowledge Engineering (2011), doi:10.1111/j.1468-0394.2010.00576.x
5. Ye, N., Li, X., Chen, Q., Masum Emran, S., Xu, M.: Probabilistic techniques for intrusion detection based on computer audit data. IEEE Trans. on Systems, Man and Cybernetics-Part A: Systems and Humans 31(4) (2001)
6. Dainotti, A., Pescape, A., Ventre, G.: Wavelet-based Detection of DoS Attacks. In: IEEE GLOBECOM, San Francisco, CA, USA (November 2006)
7. Wei, L., Ghorbani, A.: Network Anomaly Detection Based on Wavelet Analysis. EURASIP Journal on Advances in Signal Processing, Article ID 837601, 16 pages (2009), doi:10.1155/2009/837601
8. Grossman, A., Morlet, J.: Decompositions of Functions into Wavelets of Constant Shape, and Related Transforms. Mathematics and Physics: Lectures an Recent Results, L. Streit (1985)

9. Sweldens, W.: The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets. Applied and Computational Harmonic Analysis 3(15), 186–200 (1996)
10. 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)
11. BackTrack Linux, http://www.backtrack-linux.org/

# Topology Properties of Ad-Hoc Networks

Maciej Piechowiak and Sebastian Pucek

Kazimierz Wielki University, Bydgoszcz, Poland
mpiech@ukw.edu.pl, sebastian.pucek@gmail.com

**Summary.** Wireless ad-hoc networks are playing an important role in extending the implementation of traditional wireless infrastructure (cellular networks, wireless LAN, etc). Network topology planning and performance analysis are crucial challenges for network designers (i.e. routing design in ad-hoc networks is a challenge because of limited node resources). The article focuses on the network parameters determining and their influence on ad-hoc networks properties. It also proposes a new approach in ad-hoc networks modelling.

## 1 Introduction

Ad-hoc networks consists of collection of nodes placed in different geographical locations with wireless communication between them. The most distinct feature that differs them from other networks is lack of cable infrastructure – the structure is quite decentralized. Nodes in ad-hoc network can work as clients or as routers. Last few years show increased use of ad-hoc networks. They are used in military and civilian usage (on much smaller scale – used by rescue team, police or commercially by phones or computers equipped in UMTS and GPS devices). In some measurement systems nodes can represent an autonomous sensors or indicators. Ad-hoc networks can be also used to collect of sensor data for data processing for a wide range of applications such as tensor systems, air pollution monitoring, and the like. Nodes in these networks generate traffic to be forwarded to some other nodes (unicast) or a group of nodes (multicast).

Mobile ad-hoc networks (MANET) and mesh networks are closely related, but MANET also have to deal with the problems introduced by the mobility of the nodes (nodes may represent mobile devices). Similarly to the mesh networks, nodes act both as an end system (transmitting and receiving data) and as a router (allowing traffic to pass through) resulting in multi-hop routing. Networks are *in motion* – nodes are mobile and may go out of range of other nodes in the network.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 233–240.
springerlink.com 

As of today, ad-hoc networks can work in two modes: *single-hop* and *multi-hop* [2]. In *single-hop* mode, all nodes are in direct range of another node. Communication between them is possible without any external routing device. *Multi-hop* networks have the ability to be communicate and use routing device at one time. This approach improves speed of transmission and is prone to danger of losing connection. Mobility of nodes and devices multiplies problems with stability and quality of transmission.

The main goal of this article is to determine representative network parameters as average node degree, clustering coefficient and diameter, and examine their values in ad-hoc networks. Literature confirms dependencies between network topology parameters and efficiency of routing algorithms [6, 7]. The analysis of the effectiveness of the routing algorithms known to the authors and the design of the new solutions utilize the numerical simulation based on the abstract model of the existing network. These, in turn, need network models reflecting in the best adequate way the ad-hoc network. Thus new fast generator for ad-hoc networks has been proposed in the article.

The article structure is as follow: Chapter 2 describes network topology, its parameters and propagation model. Chapter 3 presents simulation study regarding authors' ad-hoc topology generator. Chapter 4 presents simulation results. The final chapter sums up the discussion.

## 2 Network Model

### 2.1 Graph Model

Let us assume that a network is represented by an undirected, connected graph $N = (V, E)$, where $V$ is a set of nodes, and $E$ is a set of links. The existence of the link $e = (u, v)$ between the node $u$ and $v$ entails the existence of the link $e' = (v, u)$ for any $u, v \in V$ (corresponding to two-way links in communication networks). With each link $e \in E$, two parameters are coupled: cost $c(e)$ and delay $d(e)$. The cost of a connection represents the usage of the link resources. $c(e)$ is then a function of the traffic volume in a given link and the capacity of the buffer needed for the traffic. A delay in the link $d(e)$ is, in turn, the sum of the delays introduced by the propagation in a link, queuing and switching in the nodes of the network.

### 2.2 Propagation Model

Ad-hoc network topologies are analyzed in many works, including [3] and [4]. These publications provide detailed analysis on modeling topologies for ad-hoc networks, methods for controlling topologies, models of mobility of nodes in networks and routing protocols in wireless ad-hoc networks. Ad-hoc networks are formed by devices that have mobile energy source with limited capacity.

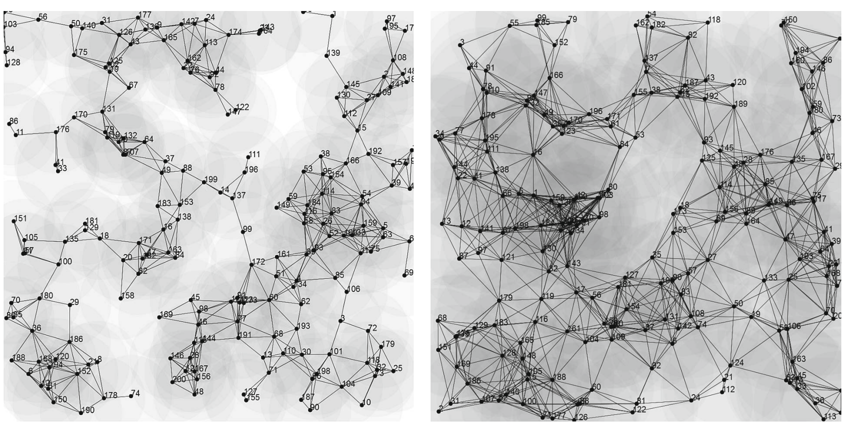
It is essential then for the energy consumption to be maintained at a possibly low level in order to prolong the time duration of autonomous operation of the device.

The adopted model of the costs of links between the devices takes into consideration energy used by the antenna system of a device. The proposed implementation assumes that network devices have isotropic radiators. The power of electromagnetic wave $P_r$ received by the antenna can be expressed by the following dependency:

$$P_r \sim \frac{P_s}{d^{\alpha}}, \qquad (1)$$

where $d$ expresses the distance between the transmitter and the receiver, and $P_s$ denotes transmitting power. If radiation propagates in vacuum, then $\alpha = 2$. However, in real environment $\alpha \in \langle 2, 6 \rangle$ [4]. In the present investigation, the value $\alpha = 3.5$ was adopted. This value was calculated as an arithmetic mean from the middle ranges of the variability of the parameter $\alpha$, published in [4] and [3]. The required power of the received electromagnetic wave $P_r$ was adopted as constant.

For simplicity, this model bases on the pathloss power law model for radio propagation. With the power law model for radio propagation, and the assumption that transmission power and receiver sensitivity for all nodes is same, the coverage area of any node is a circle with radius $r$. A node can have direct communication with all nodes that fall inside its coverage area [5].



(a) (b)

**Fig. 1** Visualization of ad-hoc networks with 200 nodes obtained using the proposed generator for $r = 100$ units (a) and $r = 150$ units (b)

### 2.3 Network Parameters

To evaluate different structures of ad-hoc networks it is important to define basic parameters describing network topology:

- *average node degree* [6]:
$$D_{av} = \frac{2k}{n} \tag{2}$$
where $n$ – number of nodes, $k$ – number of links,
- *hop diameter* [6] – the length of the longest shortest path between any two nodes; the shortest paths are computed using *hop count* metric,
- *clustering coefficient* ($\gamma_v$) of node $v$ is the proportion of links between the vertices within its neighborhood divided by the number of links that could possibly exist between them [14].
Let $\Gamma(v)$ be a neighborhood of a vertex $v$ consisting of the vertices adjacent to $v$ (not including $v$ itself). More precisely:
$$\gamma_v = \frac{|E(\Gamma(v))|}{\binom{k_v}{2}} = \frac{2|E(\Gamma(v))|}{k_v(k_v-1)}, \tag{3}$$
where $|E(\Gamma(v))|$ is the number of edges in the neighborhood of $v$ and $\binom{k_v}{2}$ is the total number of possible edges between neighbourhood nodes.
Let $V^{(1)} \subset V$ denotes the set of vertices of degree 1. Therefore [1, 12]:
$$\hat{\gamma} = \frac{1}{|V| - |V^{(1)}|} \sum_{v \in V \setminus V^{(1)}} \gamma_v. \tag{4}$$
Clustering coefficient quantifies how well connected are the neighbours of a vertex in a graph. In real networks it decreases with the decreasing value of vertex degree.

## 3 Simulation Study

Computer simulation lets turn concepts into more realistic scenarios. It allows to verify ad-hoc models and concepts without the need to implement them in hardware, yet providing a detailed insight. Therefore, authors conducted their custom-made ad-hoc generator prepared in C++, PHP [10] and SVG [11] especially for the task studies.

Generator is divided into two parts. First one is a PHP script used to convey data between user and C++ CGI applications. PHP and SVG are both used for network topology visualization. Second part consists of independent, C++ based applications that are used to quickly generate data. Dividing the generator into two parts gives much better speed than in the case the computing is done by the web server.

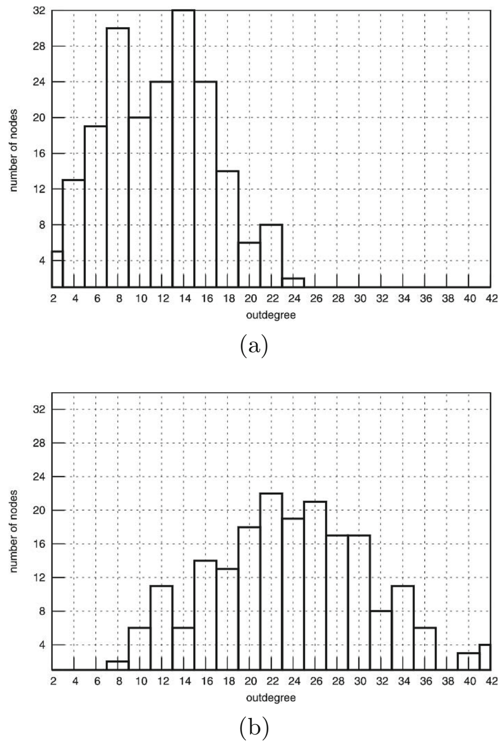



**Fig. 2** Distribution of node outdegree for ad-hoc network with $n = 200, r = 100$ (a) and $n = 200, r = 150$ (b)

Network topologies are prepared with a pseudorandom two dimensional uniform distribution generator (LCG) [9]. The simulation area is a rectangle of 1,000 by 1,000 units where nodes are deployed on a mesh with the granularity of one unit. The maximum radio range of a sensor node is set to 200 units. The proposed generator simplifies network topology model – it provides ad-hoc topologies without nodes mobility.

Figures 1a and 1b are exemplary visualizations of ad-hoc networks obtained using the proposed generator for $r = 100$ units and $r = 150$ respectively. The second network has higher average node degree.

## 4 Simulation Results

In the first phase of the experiment (Fig. 2) distribution of node outdegree for ad-hoc network with $n = 200$ (histograms) were examined for the networks topologies presented on Fig 1. The range $r$ (representing transmission power

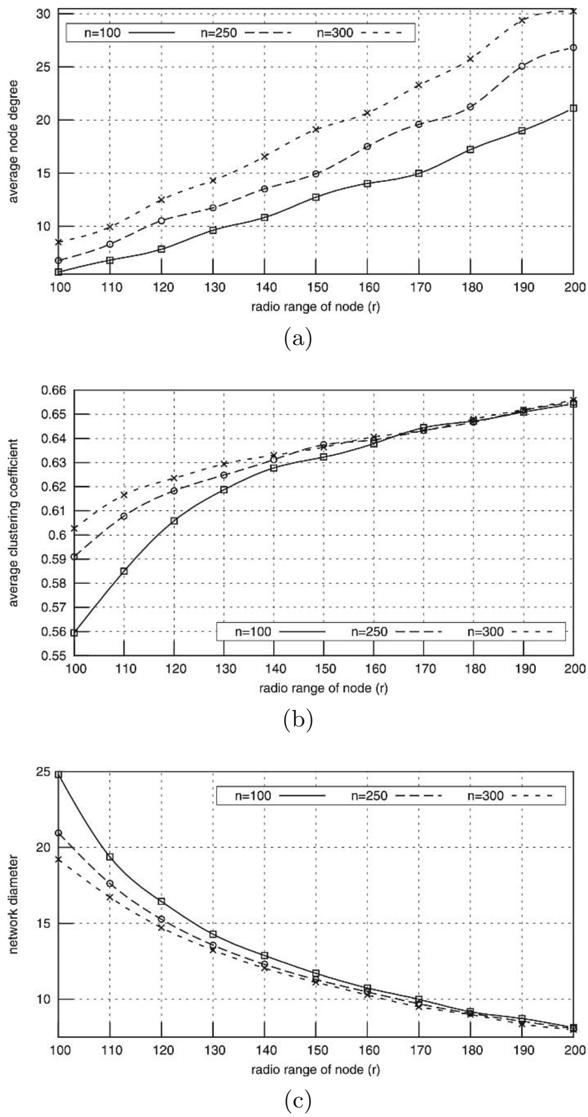



**Fig. 3** Average node degree (a), average clustering coefficient (b) and network diameter (c) versus to the radio range of node ($r$)

level) has significant influence on outdegree distribution. For $r = 100$ network represented by undirected graph contain *leaves* – nodes with outdegree equal 2 (5 nodes). Small value of node outdegree is also noticeable (43% nodes have outdegree from the range of 2 to 10). Node outdegree increases with the increasing value of $r$ (for $r = 150$ network has no leaves and there are 7 nodes with outdegree 42 and 43). Further increasing the radio range constructs full-mesh networks (histogram with one, high outdegree peak).

In the second phase of the experiment average node degree ($D_{av}$), average clustering coefficient ($\hat{\gamma}$) and hop-diameter were examined in relation to the radio range ($r$). Increasing value of $r$ (Fig. 3a) results in increasing average node degree (270% increment of $D_{av}$ is observable for $r$ from the range of 100 to 200). Average clustering coefficient ($\hat{\gamma}$) is differentiated when radio range $r = 100$ and networks have different number of nodes $n$ (Fig. 3b). For $r > 170$ average clustering coefficient has the same values independent of number of network nodes $n$. Increasing value of $r$ results in decreasing hop-diameter (Fig. 3c). Hop-diameter value is biggest for $n = 100$ and $r = 100$ (small networks with small radio ranges).

The simulation outcomes presented in the paper are the average results computed for many independent simulation iterations (Fig. 1). The values of the simulations have 95% confidence intervals calculated after the t-Student distribution. The confidence intervals are so small that, for most of the cases, they are within measurement points shown in the figures. For the sake of readability and convenience they are not shown in the graphs.

## 5 Conclusions and Future Work

The article defines representative network parameters as average node degree, clustering coefficient and diameter, and examine their values in ad-hoc networks. Thus new fast generator for ad-hoc networks has been proposed in the article.

Previous authors' works show strong influence between basic network parameters and results of routing algorithms [7, 8]. There is a need to confirm these tendencies in ad-hoc networks.

For the purposes of the study, it is assumed that future, far more advanced, devices will have the capability of precise fine tuning of the transmitting power level to that required by the receiver. It is an interesting of further research work to extend proposed generator to model the direction and transmitting power level of each node.

## References

1. Bu, T., Towsley, D.: On distinguishing between Internet power law topology generators. In: Proceedings of INFOCOM (2002)
2. Conti, M., Crowcroft, J., Passarella, A.: Multi hop Ad hoc Networks from Theory to Reality. EURASIP J. Wireless Communications and Networking, special issue (2007)
3. Rajaraman, R.: Topology Control and Routing in Ad Hoc Networks: A Survey. ACM SIGACT News 30, 60–73 (2002)

4. Santi, P.: Topology Control in Wireless Ad Hoc and Sensor Networks. ACM Computing Surveys 37, 164–194 (2005)
5. Grover, P., Gupta, N., Kumar, R.: Calculation of Inference in Ad-Hoc Network. J. Theoretical and Applied Information Technology 16(2), 105–109 (2010)
6. Zegura, E.W., Calvert, K.L., Bhattacharjee, S.: How to Model an Internetwork. In: IEEE INFOCOM 1996 (1996)
7. Piechowiak, M., Zwierzykowski, P.: The Influence of Network Topology on the Efficiency of Multicast Heuristic Algorithms. In: Proceedings of The 5th International Symposium – Communication Systems, Networks and Digital Signal Processing, CSNDSP 2006, pp. 115–119 (2006)
8. Piechowiak, M., Zwierzykowski, P.: Performance of Fast Multicast Algorithms in Real Networks. In: Proceedings of International Conference on Computer as a Tool EUROCON 2007, pp. 956–961 (2007)
9. Generator Services Project, http://sftweb.cern.ch/generators/
10. PHP: Hypertext Preprocessor, http://sftweb.cern.ch/generators/
11. Scalable Vector Graphics (SVG), http://www.w3.org/Graphics/SVG/
12. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On Power-Law Relationships of the Internet Topology. ACM Computer Communication Review, 111–122 (1999)
13. Zegura, E.W., Donahoo, M.J., Calvert. K.L.: A Quantitative Comparison of Graph-based Models for Internet Topology. IEEE/ACM Transactions on Networking (1997)
14. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature 12(393), 440–442 (1998)

# Computational Complexity and Bandwidth Resource Analysis of NLM Interleaver in IDMA System

Brahim Akbil[1], Guillaume Ferre[2], and Driss Aboutajdine[1]

[1] LRIT URAC29, Mohammed V-Agdal University, Morocco
akbil.brahim@ieee.org, aboutaj@fsr.ac.ma
[2] University of Bordeaux 1, IMS Laboratory CNRS UMR 5218, Signal & Image Team
guillaume.ferre@ims-bordeaux.fr

**Summary.** In this paper, we develop a new one-dimensional chaotic map, completely based on the logistic map: "New Logistic Map (NLM)"; this new map expand the definition domain of logistic map and their utilization. The simulation results and theoretical analysis of chaotic dynamical behaviours show that the NLM is a chaotic system and has an ideal distribution. It is also characterized by its wide range of definition, and depends on the number of the desired real values.We have applied the NLM to generate the interleaver matrix for Interleave Division Multiple Access (IDMA): NLMInterleaver. Our design easy to generate, and requires a low memory bandwidth consumption (low amount of data can travel over a communications channel), and low computational complexity compared to the literature algorithms.

## 1 Introduction

Interleave Division Multiple Access is one of the main candidate multiuser detection techniques based on the interleaver matrix for next generation mobile communication systems. This technique inherits the CDMA advantages with one spreading sequence for all users, and proves to be a particular case of the CI-CDMA, relies on interleaving to distinguish and separate signals from different users [1]. The interleaver/deinterleaver is the main component in the IDMA transmitter and receiver; however the choice of better interleaver has a positive influence on the IDMA performances. The main criteria to choose a good interleaver is that each two interleavers out of a set of interleavers should not "collide" [2], including minimal bandwidth consumption to exchange the information about interleaver matrix between transmitter and receiver, minimal memory required to store the interleaver matrix, and ease to generate it. However, in the standard case the transmitter needs to transmit the entire interleaver matrix consisting of interleaving pattern of the users to the receiver,

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 241–251.
springerlink.com 

thus the large value of the interleaver size requires a high memory to store the interleaver matrix, more bandwidth and resources will be needed and the implementation complexity will be high.

Several designs of interleaver were introduced and studied in the literature, such as the Orthogonal Interleaver (OI), Random Interleaver(RI), Pseudorandom Interleaver [2], Shifting Interleavers (SI) [3], Nested Interleaver (NI) [4], Deterministic Interleaver (DI) [5] and others [6, 7]. The orthogonal interleaver is generated by using the orthogonal binary sequences and its main limit is that the number of simultaneous users is limited by the spreading sequence length. Since in the family of random interleavers, the generator algorithm is not described, the exchange of the interleaver matrix between the transmitters and the receivers is necessary thus a large bandwidth will be consumed; consequently, increasing the length of the interleaver matrix will increase the consumption of the bandwidth. In the nested interleaver, the first base interleaver is pseudorandomly generated and the next interleavers are generated by reindexing operation. The OI, RI and NI are useless to use in next-generation due to the fact that the first interleavers are randomly generated, and the computational complexity is proportional to the number of users. In addition, the bases interleavers matrix have to be transmitted to the receiver. To easily produce an interleaver with deterministic behavior, another design, called Deterministic Interleavers (DI), was proposed in [8]; it is firstly used for Turbo Code/Decode, and adapted in [5] for IDMA systems. This interleaver design ameliorates the bandwidth consumption problem. Moreover, the disadvantage of DI is the requirement of a large memory in the generation step; as a result as they still use a Look-Up Table (LUT) where basic mapping function has to be stored after its computation.

In this context, we introduce a new one-dimensional chaotic map in this paper, which we name "New Logistic Map (NLM)", we also study its dynamical properties including invariance, phase space and bifurcation diagram. We apply these properties to propose a new design that will allow us to construct a set of interleavers matrix for different users in IDMA systems. The novel advantages of our interleaver design are the simplicity to generate it, least memory requirement, and minimal consumption of bandwidth.

The content of the paper is organized as follows. In section 2, we give an introduction to our new map, and analyze the chaotic dynamical behaviour in this map. In Section 3, we revisit the IDMA system given by Li Ping. In Section 4, we present the details to construct NLM Interleaver. Section 5 presents the simulation results, and finally section 6 gives conclusions.

## 2 New Logistic Map

The logistic map [9] is one of the simplest and most transparent discrete dynamical systems exhibiting the order to chaos transition. The interest of this map is due to its important characteristics, such as deterministic, sensitive to initial conditions, motion ergodic and embedded with an infinite number of unstable periodic orbits. Mathematically, the logistic map is written as

$$X_{n+1} = \lambda X_n (1 - X_n), \tag{1}$$

Where $X_n \in [01]$, and $n$ is the generation number. Contrary to the randomsequence, the one generated by logisticmap is predictable and hence reproducible from the initial state $X_0$ and a deterministic algorithm. A sequence generated by the logistic map consists of a Cantor set of points in [01] whose orbits stay in [01]. However, many applications use the sequences of $N$ random integers in $[0N]$ such as the interleaver/deinterleaver in IDMA systems. Let us consider a new system given by:

$$X_{n+1} = \lambda X_n \left(1 - \frac{X_n}{N}\right), \tag{2}$$

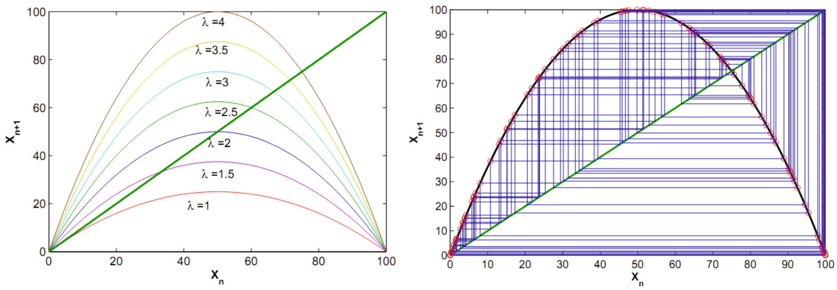
Where $X_n \in [0N]$ and $N$ represent the interleaver length. This system must be chaotic and there must be many interesting complex dynamical behaviors. To check the numerically chaotic behavior in this system, we analyzed the phase space in the Cartesian plane in the term of the parameter $\lambda$ and the bifurcation diagrams.

### 2.1 Phase Space

Fig. 1a shows the phase space in the Cartesian plane of (2) for different values of $\lambda$. If $0 < \lambda < 1$, the equilibrium at $X_0$ is stable and if $1 < \lambda \leq 3$, the system (2) is stabilized in equilibrium $\frac{\lambda-1}{\lambda}$. At $3 < \lambda \leq 3.54$ we have a violent behavior, in this case the system oscillates on a cycle of 4 periods without reaching the equilibrium. At $\lambda = 3.57$ chaos occurs; the system never settles to a fixed period. For $3.57 < \lambda \leq 4$, the system is aperiodic (Fig. 1b), since the appearance of the chaotic behavior.

### 2.2 Bifurcation Diagrams

Another step to evaluate the chaotic behavior of the NLM system is to analyze a visual summary of the succession of period-doubling produced as the control parameter $\lambda$ when it is varied. This process is called bifurcation diagrams, which are tools commonly used in nonlinear dynamics. Fig. 2 shows the bifurcation diagram of the NLM function for a sequence of $N = 100$ samples, and $0 < \lambda \leq 4$. We see that for $\lambda \leq 1$, we have only one point attractor for our system, this point is zero. For $1 < \lambda \leq 3$, we still have one point attractor, but the value of $X$ increases when $\lambda$ increases. At $3 < \lambda \leq 3.57$ the entire bifurcation is recreated on a smaller scale, and for $\lambda > 3.58$ the system is completely chaotic.




(a) (b)

**Fig. 1** Chaotic behavior in NLM for (a): different values of $\lambda$, (b): $\lambda = 4$ and $X_0 = 1.2$

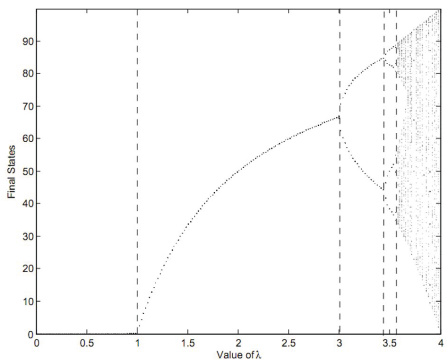



**Fig. 2** The bifurcation diagrams for NLM

## 3 IDMA System

In Fig. 3, the transmitter and the receiver iterative MultiUser Detector (MUD) structure of an IDMA system with $U$ simultaneous users is presented. The input data sequence $d_k^u$ of $u^{th}$ user is spread based on a low-rate code $C$ ($C$ can be either the same or different for different users) to construct a sequence $b^u$. Then $b^u = [b^u(1), b^u(2), \ldots, b^u(N)]_T$ - where $N$ is the frame length- is permuted by an interleaver $\pi^u$ to product a signal $f^u = [f^u(1), f^u(2), \ldots, f^u(J)]_T$ where the basic elements in $b_j^u = d_k^u C^u$ called "chips". The different users are distinguished mainly by their interleavers $\pi^u$. The received signal from $U$ users can be written as follow:

$$R_j = \sum_{u=1}^{U} f_j^u h^u + \mathscr{N}_j ; j = 1, 2, \ldots \tag{3}$$

where $\mathcal{N}_j$ is the $j^{th}$ sample of an Additive White Gaussian Noise (AWGN) with variance $\sigma^2 = \frac{N_0}{2}$ end zero-mean, and $h^u$ is the channel coefficient for $u^{th}$ user. As illustrated in Fig. 3, the MUD receiver consists of an Elementary Signal Estimator (ESE) and $U$ single user a posteriori probability decoders (APP-DEC). The ESE computes the extrinsic Log-Likelihood Ratio (LLR) about $f_j^u$ given by:

$$e_{ese}^u = \ln \left[ \frac{P(R_j|f_j^u = +1}{P(R_j|f_j^u = -1} \right] \quad (4)$$

In our system, the DEC consists of $U$ local APP decoders and each decoder is the dispreading and spreading operation. After appropriate deinterleaving, the $u^{th}$ local APP decoder performs an APP decoding of $C$ for user $u$. The extrinsic LLRs are used to update the mean and the variance of the interference and noise components in the received signal $R_j$.

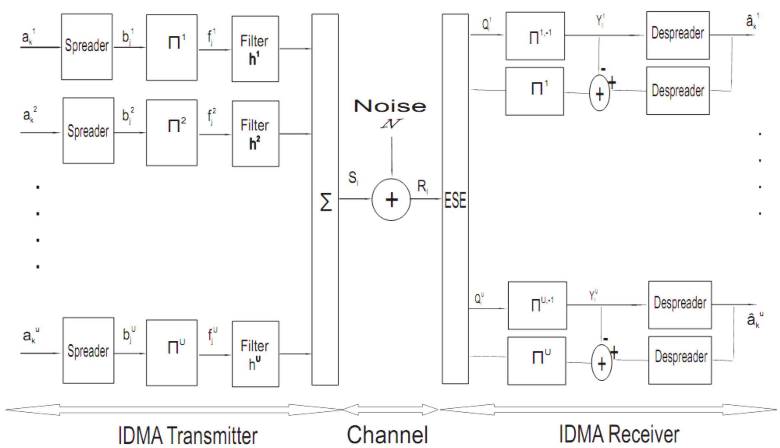



**Fig. 3** IDMA system structure with $U$ users

To begin the next iteration of the MUD process, the extrinsic information output of DEC is then respread, and reinterleaved before feed back to the ESE. The algorithm to estimate the mean and the variance in ESE, and DEC of every chip in every $f_j^u$ is detailed in [10].

## 4 Design Steps of the NLM Interleaver

The graph of NLM function in the Cartesian plane is a parabola, that illustrates the phase space correspond to the states occupied by the system during its evolution; this characteristic is the basic idea of our method to generate interleavers chaotic. The system starts with a first value $X_0$ of $X$ called the initial state and then calculates a sequence $\{X_1, X_2, \ldots, X_N\}$ occupied by the system using our system $X_{n+1} = \lambda X_n \left(1 - \frac{X_n}{N}\right)$; these states form a set of real sequences between 0 and $N$. The transition to the integer sequences is realized by maximizing, or by using a floor to round the elements of $X_i$ to the nearest integers. Finally to obtain the interleaver vector, we eliminate redundant elements from the vector found. Starting from an initial value $X_0^u$ we obtain -by iterating the NLM given in (2)- the interleaver matrix for $u^{th}$ user. We construct an initial value for the next user by adding a footstep $\zeta$ to the $X_0^u$. In our method, it is not necessary to transmit large amount of interleaver matrix to the receiver side, it only transmits the initial value $X_0^u$ for each user. All users have the same complexity to generate these interleavers. The following is a list of our algorithm -for each user- based on the NLM equation:

Initialization:

- $\lambda = 4$ The bifurcation parameter, $N$: Interleaver length,$U$: The number of users, $i = 0$ and $n = 0$.
- $X_i^u$: The initial state of $u^{th}$ user $(0 < X_i^u < N)$, $\zeta$: The footstep between the initial states, $u = 1$.
- $Y_0^u = \int(X_0^u)$: The first element interleaver matrix of $u^{th}$ user $(\pi^u \equiv Y_0^u)$.

Main operations: the value of $n$ does not exceed $N$:

1. Calculate $X_{i+1}^u$ from the value of $X_i^u$.
2. Use floor to round the elements of $X_{i+1}^u$ to the nearest integers $(Y_{i+1}^u = \int(X_{i+1}^u)$.
3. Check, if the element $Y_{i+1}^u$ exists in the set $\pi^u$, increment $i$ by 1 and repeat the second and the third steps in the main operations, otherwise store the element $Y_{i+1}^u$ in the set $\pi^u(\pi^u \equiv \pi^u \bigcup Y_{i+1}^u)$ and increment $n$ by 1.

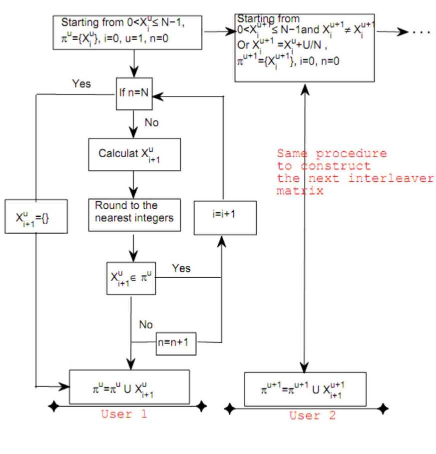



**Fig. 4** Interleaver construction based on the NLM

It is important to estimate the initial state, the perfect footstep between the initial states and the technique used to pass from a real to integer sequences; therefore, these parameters are used to find good interleavers, thus minimize the Inter User Interference (IUI).

# 5 Performance of NLM Interleaver

## 5.1 Computational Complexity

The complexity of interleavers and deinterleavers generation at the transmitter and receiver side is a major concern in case of higher user count [2]. In this section we compare the computational complexity to generate the interleaver matrix in different designs.

The algorithm to generate the orthogonal interleavers is more detailed in [2]. The main limit of this interleaver designs is that the number of simultaneous users is limited by the spreading sequence length. The computational complexity of an orthogonal interleaver equals the computational complexity to generate a PN sequence using a linear feedback shift register (LFSR) with a primitive polynomial and the complexity to construct the interleaver matrix from PN sequences.

The schifting interleavers are also based on the PN-sequences generate with primitive polynomial to generate a master interleaver. To construct the other interleaver, a circular shifting of master interleaver is used. The nested interleaver is generated by using composition map of a single interleaver. The second nested interleaver is generated by the permutation of the image of the first interleaver by itself. The next interleaver is also generated by the permutation of the image of the second interleaver by using the first interleaver. This operation is repeated to obtain the next interleavers. However, the operation to generate the next interleavers is slower after the generation of the first interleaver; thus the computational complexity of nested interleavers is approximately equal to that of orthogonal interleavers.

Table 1 shows the comparison of computational complexity among the NLM interleaver design and the reference designs. This complexity is calculated by the number of cycles needed to generate the interleaver matrix versus the number of users. As expected, Table 2 shows that the computational complexity increases with the number of users for orthogonal, shifting and nested designs while it is not the case for NLM interleaver designs. Therefore, the computational complexity in NLM interleavers is fixed and independent of the number of users. $\Im_{nt} = \int(\frac{N}{U})$ returns the greatest integer that is not larger than $\frac{N}{U}$.

## 5.2 Bandwidth Resource Required

Lack of bandwidth resource is a vital issue in communication systems. This problem becomes worse when the user number increases. In IDMA system,

**Table 1** The computational complexity to generate the first interleaver

| | Orthogonal | Shifting | Nested | NLM |
|---|---|---|---|---|
| Complexity | $O(N)+O(N^2)$ | $O(N^2)$ | $O(\log_2(N))$ | $O(N^2)$ |

**Table 2** The number of cycles needed to generate the $u^{th}$ interleaver

| $u$ | Orthogonal | Shifting | Nested | NLM |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 16 | 16 | $16\times\Im_{nt}$ | 16 | 1 |
| 200 | 200 | $200\times\Im_{nt}$ | 200 | 1 |

the number of simultaneous users is a very important factor. However, a great number of researchers are interested in increasing the maximum available number with a minimal consumption of bandwidth. Another problem in IDMA system is that the transmitter and receiver must hold the same interleaver matrix. In most of the studied algorithms, the transmitter needs to transmit the interleaver matrix consisting of interleaving pattern of the users to the receiver; so the greater the size of the interleaver, the more bandwidth and resources are used.

To evaluate the bandwidth consumption, we propose here a brief overview of the initialization parameters required to generate the interleavers and the number of bits occupied by users in the different designs. The random interleavers generated randomly and independently and the selection of user specific interleavers is also performed randomly for all the users. These randomly selected interleavers, at transmitter end, are required to be sent to receiver end for proper decoding of user-specific data. However, a high bandwidth and memory requirement are needed. The orthogonal interleaver is generated by using the orthogonal binary sequences and the pseudonoise (PN) sequences generated by a linear feedback shift register defined by a primitive polynomial of degree $m$ over the Galois field GF(2). The amount of exchanged information between transmitter and receiver are a primitive polynomial of degree $m$, where $2^m > N$, and the initial state of the linear feedback shift register with $m$ bits.

In the case of nested interleavers, the generation process begins by initializing the state of the linear feedback shift register with a primitive polynomial and a pseudo random interleaver generated by using the bits represents the coefficients of the primitive polynomial. It is necessary to store the "seed" of these interleavers in the mobile stations. According to the first interleaver, the $(u+1)^{th}$ interleaver can be obtained with permutation $\pi^u o \pi^1$, with $u = 1, 2, \ldots, U$. In order to agree upon a nested interleaver between the transmitter and the receiver, and then generate it. They are necessary to exchange a primitive polynomial, the initial state of the linear feedback shift register and a pseudo random interleaver.

**Table 3** The number of bits required for different interleaver designs

| | $\mathbf{b}_M$ | $\mathbf{B}_M$ |
|---|---|---|
| Random | $\log_2(N1) \times (N+2)$ | $U \times \log_2(N-1) \times (N+2)$ |
| Orthogonal | $2 \times m + 3 \times \log_2(N-1)$ | $2 \times m + 3 \times \log_2(N-1)$ |
| Shifting | $\log_2(N-1) \times N + 2\times$ $\times m + 3 \times \log_2(N-1)$ | $\log_2(N-1) \times N + 2\times$ $\times m + 3 \times \log_2(N-1)$ |
| Nested | $2 \times m + \times \log_2(N-1)$ | $2 \times m + 3 \times \log_2(N-1)$ |
| NLM | $4 \times \log_2(N-1)$ | $4 \times \log_2(N-1)$ |

**Table 4** The number of bits required $\mathbf{B}_M$ as a function of $U$ for different interleaver designs with $N = 64$

| U | Random | Orthogonal | Shifting | Nested | NLM |
|---|---|---|---|---|---|
| 1 | 396 | 30 | 414 | 30 | 24 |
| 64 | 24588 | 30 | 414 | 30 | 24 |
| 128 | 49164 | - | 414 | 30 | 24 |
| 256 | 98316 | - | 414 | - | 24 |

In NLM interleaver, a single information exchange between the transmitter and the receiver is only the initial value $X_0^1$. The interleaver matrix for first user is generated based on the formula (2) , and the interleaver for $u^{th}$ user is generated by initializing the formula (2) by $X_0^u = X_0^1 + u \times \zeta$.

Let $\mathbf{b}_M$ the portion length of the IDMA frame reserved to exchange the information between transmitter and receiver to agree upon an interleaver and then generate it for each user and $\mathbf{B}_M$ the bandwidth consumption to generate the interleavers matrix for all users.

Let us consider a primitive polynomial $g(x) = x^6 + x + 1$ of degree $m = 6$ and [1 0 1 1 0 0] the initial state of the linear feedback shift register used to generate an PN-sequence with $g(x)$. In Table 3, we show the expression of the number of bits to be transmitted to the receiver to generate the same interleaver matrix as is generated in the transmitter, while Table 4 shows an example with $N = 64$ as a function of the number of simultaneous users $U$.

subsectionBER performance results The next goal of this paper is to evaluate the quality of transmission using bit error rate (BER) analysis of IDMA system for proposed NLM interleaver along with RI. For simplicity, we evaluate an IDMA system with BPSK signaling in single path AWGN channel. The block size of data bits for each user is 512 and the used spreading code is $C = [1, 1, 1, \ldots, 1]$ for all users with the length $S = 32$. The number of iterations is set to be 10 in each case.

In Fig. 5, the comparative simulation of NLMI and RI is showed for 512 bits of data length with user number $U = 1, 32, 64, 96, 110$. The simulation result demonstrates that the NLM interleaver achieves similar BER performance compared to other designs, in case of similar simulation conditions.

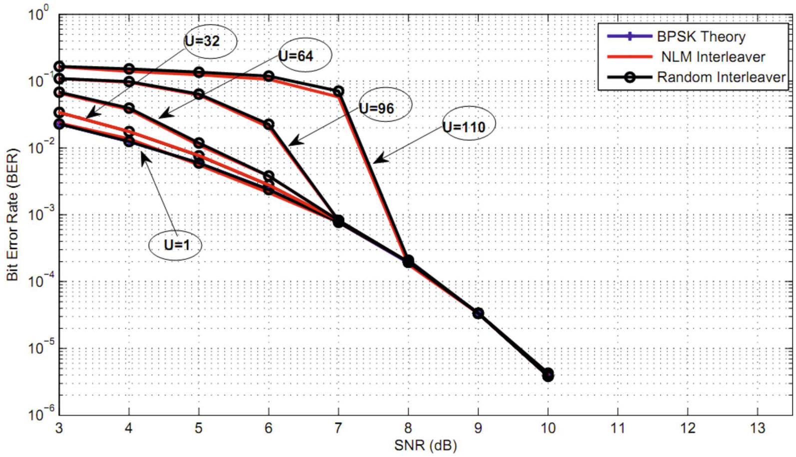



**Fig. 5** Performance comparison of NLMI and RI in IDMA system with $N = 512$ using various number of users $U$

## 6 Conclusion

We claim two contributions of this work. First we have developed a new chaotic map equation called New Logistic Map (NLM), and we have analyzed its chaotic dynamical behaviors in detail. In the second main contribution, we have outlined the role of interleaver design in IDMA system and we have proposed a novel method to generate good and multiple interleavers, regardless of the number of simultaneous users and the interleaver length. Our proposed interleaver design is based on the NLM and called NLM interleaver. The simulations in Section 5 show that the NLM interleavers can achieve the same BER performances compared to random interleavers, shifting interleavers and nested interleavers. NLM interleaver has also many advantages such as less resource consumption, less required memory, minimum implementation complexity and easy to generate it, over random or any other interleavers.

## References

1. Ping, L., Wu, K.Y., Liu, L., Leung, W.K.: A Simple Unified Approach to Nearly Optimal Multiuser Detection and Space-Time Coding. In: ITW 2002, Bangalore, India, pp. 20–25 (2002)
2. Pupeza, I., Kavcic, A., Ping, L.: Efficient generation of interleavers for IDMA. In: IEEE International Conference on Communications, ICC 2006, vol. 4, pp. 1508–1513 (2006)
3. Zhang, C., Hu, J.: The shifting interleaver design based on PN sequence for IDMA systems. In: 2007 International Conference on Future Generation Communication and Networking, FGCN 2007, Korea (2007)

4. Wu, H., Ping, L., Perotti, A.: User-specific chip-level interleaver design for IDMA systems. IEE Electron. Lett. 42(4), 233–234 (2006)
5. Tseng, T.-M.: IDMA Based on Deterministic Interleavers. International Journal of Communications, Network and System Sciences 3(1), 94–97 (2010)
6. Zhang, C., Hu, J.: 2-dimension interleaver design for IDMA systems. In: Proc. IEEE International Conference on Circuits and Systems for Communications, ICCSC 2008, pp. 372–376 (2008)
7. Han, L., Jin, M., Song, E.: Matrix Cyclic Shifting Based Interleaver Design for System IDMA. In: Proc. 5th International Conference on Wireless Communications, Networking and Mobile Computing, WiCom 2009, pp. 1–4 (2009)
8. Arif, M., Sheikh, N.M., Sheikh, A.U.H.: A Novel Design of Deterministic Interleaver for Turbo Codes. In: Proc. Int. Conference on Electrical Engineering (2007)
9. May, R.: Simple mathematical models with very complicated dynamics. Nature 261, 458–467 (1977)
10. Ping, L., Liu, L., Wu, K., Leung, W.K.: Interleave-Division Multiple-Access. IEEE Transactions on Wireless Communication 5(4), 938–947 (2006)

# Support Vector Machine Classification in a Device-Free Passive Localisation (DfPL) Scenario

Gabriel Deak[1], Kevin Curran[1], Joan Condell[1], Daniel Deak[2], and Piotr Kiedrowski[3]

[1] Intelligent System Research Centre, Northland Road, Derry, N. Ireland, UK
`Deak-G@email.ulster.ac.uk`, `{KJ.Curran,J.Condell}@ulster.ac.uk`
[2] S.C. Centrul de Calcul Info98 S.A., 2 Timisoara Street, 332015, Petrosani, Romania
[3] Institute of Telecommunication, University of Technology and Life Science, ul. Kaliskiego 7, 85-789 Bydgoszcz, Poland

**Summary.** The holy grail of tracking people indoors is being able to locate them when they are not carrying any wireless tracking devices. The aim is to be able to track people just through their physical body interfering with a standard wireless network that would be in most peoples home. The human body contains about 70% water which attenuates the wireless signal reacting as an absorber. The changes in the signal along with prior fingerprinting of a physical location allow identification of a person's location. This paper is focused on taking the principle of Device-free Passive Localisation (DfPL) and applying it to be able to actually distinguish if there is more than one person in the environment. In order to solve this problem, we tested a Support Vector Machine (SVM) classifier with kernel functions such as Linear, Quadratic, Polynomial, Gaussian Radial Basis Function (RBF) and Multilayer Perceptron (MLP) in order to detect movement based on changes in the wireless signal strength.

## 1 Introduction

Indoor location estimation is a crucial component in many applications. Location estimation is important for many scenarios such as asset tracking, health care, location based network access, games, manufacturing, government, logistics, industry, shopping, security, tour guides, and conference guides. Various localisation systems that can estimate the position of a person or object exist. One can select the system which offers the accuracy and precision required for a specific application.

Indoor localisation systems can be classified into active and passive systems. Location tracking techniques for active localisation require the tracked people to participate actively. The second class known as passive localisation

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 253–260.
springerlink.com 

is based on monitoring changes of characteristics dependent on peoples presence in an environment. By participating actively, we mean that a person carries an electronic device which sends information to a positioning system helping it to infer that person's position. In some cases the electronic devices can also process recorded data and send the results for further processing to an application server running the localisation algorithm. In the passive localisation case, the position is estimated based on the variance of a measured signal or video process. Thus the tracked person is not carrying any electronic devices to infer the user's position.

This work is focused on solving an extremely difficult task that is multi-occupancy detection in a passive localisation scenario. Thus the following sections will analyse one of the techniques used to deploy indoor passive localisation systems. Various DfPL systems will be presented as an introduction to indoor passive localisation. Various techniques such as Ultra-wideband (UWB), Physical Contact, Differential Air Pressure, Computer Vision, and Device-free Passive Localisation (DfPL) have been used in indoor passive localisation.

Ultra-wideband (UWB) is one of the first techniques used to deploy passive localisation systems [1]. Through-the-wall surveillance or through-wall imaging (TWI) are used to denote UWB passive systems [2, 3]. This technique has been recently used for both static and motion detection. UWB passive localisation is considered to be an extension to a technique called radio tomographic imaging due to its similarity to the medical tomographic imaging. Through-wall imaging refers to the ability of detecting and monitoring objects or people through buildings walls. This can be very useful to law enforcement agencies and can have many applications in military and civil scenarios [4]. UWB has the advantage of being able to penetrate walls. Various implementations of UWB technique have been proposed. A UWB system has the following two main components: transmitters and receivers. Short pulses are sent by a pulse generator via a horn antenna [5]. The receivers wait and monitor echoes from various objects or people.

TileTrack represents a low cost two-dimensional location estimation system based on physical contact [6]. Changes in the capacitance between transmitting and receiving electrodes (plate electrodes or wire electrodes) are monitored. The system is based on 9 floor tiles with one transmitting electrode for each tile. Each tile is 60 cm by 60 cm square-shaped made from thick chipboard with thin steel coating. The prototype used to deploy the TileTrack technique has a square tracking area with a size of 3 $\times$ 3 tiles.

AirBus estimates location based on indoors airflow disruption caused by human movement [7]. An air pressure sensor is placed within the central heating, ventilation, and air conditioning (HVAC) unit. The sensor detects pressure variations. AirBus can correctly identify an open or closed door 80% of the cases with HVAC in operation and 68% with HVAC unit switched off.

Computer vision can be considered as a DfPL system because the tracked people are not carrying any electronic devices or tags [8]. The EasyLiving project [9] is a computer vision based system which aims to transform any environment in a smart environment dependent on location information. Possible applications include switching on/off devices near to the users location, monitoring peoples behaviour and many others. The system architecture consists of three PCs (Personal Computers) and two sets of colour cameras. Each camera is connected to one PC, while the third PC is used for running the person tracker algorithms. Video processing algorithms are used to separate and track people. The system was tested with a maximum of three people in the environment. The possibility of obstructions depends on the behaviour and the number of persons.

The Device-free Passive Localisation (DfPL) [10, 11] is based on monitoring the variances of the signal strength in a wireless network. The human body contains about 70% water and it is known that waters resonance frequency is 2.4 GHz. The frequency of the most common wireless networks is 2.4 GHZ, thus the human body behaves as an absorber attenuating the wireless signal [2, 4, 12, 13, 14, 15]. This technique is the focus of our research and the remainder of the paper is based on DfPL using Wireless Sensor Networks (WSNs).

The paper is organised as follows: Section 2 introduces Support Vector Machine Classification with various kernel functions, Section 3 presents the test bed and motion detection technique using the classifier introduced in Section 2. Section 4 concludes the paper.

## 2 Suport Vector Machine (SVM) Classification

SVM is a supervised learning method for data analysis, pattern recognition, classification and regression analysis. SVM uses training vectors, pairs of inputs-outputs, to build a model that is used afterwards to predict classes that new data belongs to. For our tests we used two Matlab functions *svmtrain* and *svmclassify* defined by:

$$
\begin{aligned}
&SVMStruct = svmtrain(Training, Targets, Name, Value)\\
&PredClass = svmclassify(SVMStruct, TestData)
\end{aligned}
\tag{1}
$$

*Training* and *Targets* represent the input-output pairs used for training. *Name-Value* pair specify optional arguments. *svmclassify* uses the obtained model to classify new data. One can find more details about SVM classification in Matlab including all the optional parameters in [16]. We tested various kernel functions as follows:

$$
\begin{aligned}
SVMStruct =& svmtrain(Training, Targets,\\
&'kernel\_function', 'polynomial')
\end{aligned}
\tag{2}
$$

where *'kernel_functions'* is an optional argument and *'polynomial'* represents the type of kernel function used for training. The training process can use kernel functions such as Linear, Quadratic, Polynomial, Gaussian RBF and MLP. *SVMStruct* represent the model obtained after training. This is a structure containing information about the trained SVM classifier. A field of interest in this structure/model is *GroupNames* which returns the predicted classes for the data represented by *TestData*, a parameter of *svmclassify* function.

Section 3 of this paper presents results obtained with SVM classifier in a DfPL scenario.

## 3 Test Bed and Results

This section presents the experiment we conducted in order to detect motion in a DfPL scenario. First, the test bed shown in Figure 1 will be described and then Support Vector Machine classifier will be used to analyse/classify motion. Finally, we compare the classification errors for various kernel functions in Table 1.

The project focuses on deploying a DfPL system on top of a Wireless Sensor Network (WSN). The first step towards implementing such a system is filtering data using a selected smoothing algorithm, Savitzky-Golay smoothing filter in this case. The filtered data is fed to a SVM classifier in order to detect movement. The next step is using classified data, timestamps and links affected as parameters in a decision making algorithm that will compute a person's location or return the number of people detected. The project aims towards multi-occupancy detection in a DfPL scenario.

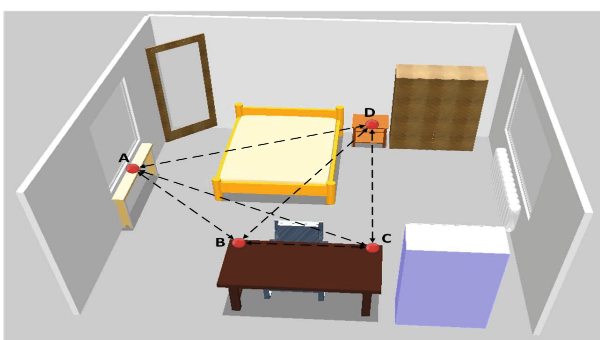



**Fig. 1** The test bed with bidirectional link selected

We have collected the data in a room of size 3.6m by 3.4m. A Wireless Sensor Network (WSN) based on four IEEE 802.14.5 Java Sun SPOT nodes and a base station was deployed in the environment. The data was recorded using a single thread collection over a period of approximately two hours. The data sets were obtained by observing Received Signal Strength Indicator (RSSI) levels when movements took place. We use two data sets containing 800 values (see Figure 2) selected from the two hours recordings. The first data set represents the training data while the second one is the test data. The nodes are broadcasting messages every 200 ms. When the messages are received, Received Signal Strength Indicator (RSSI) is added and then the messages are forwarded to the base station. However, working with a single collection thread can cause delays as the base station collects data from one node at a time.

In the case of four nodes, considering that we collect data from 12 links, we experienced delays in the collection speed. However the delays were not large enough to affect our tests. For larger test beds, multiple collection threads or more than one base station will improve the collection speed.

We have selected one bidirectional link between nodes A and B as shown in Figure 1. Both links are considered to be independent. Figure 2 shows the raw data collected from the selected links. We do not use data collected from all 12 links as the scope of the paper is to classify motion on two selected links (bidirectional communication between nodes A and B). SVM classification will perform in a similar manner on any selected link.

Both data sets are smoothed in order to filter noise. The derivative of the signal is used to normalise the data. Figure 3 shows the smoothing and derivative on one of the links. Data from the second link is processed in a similar manner. It is necessary to normalise the data in order to train and use classifiers.

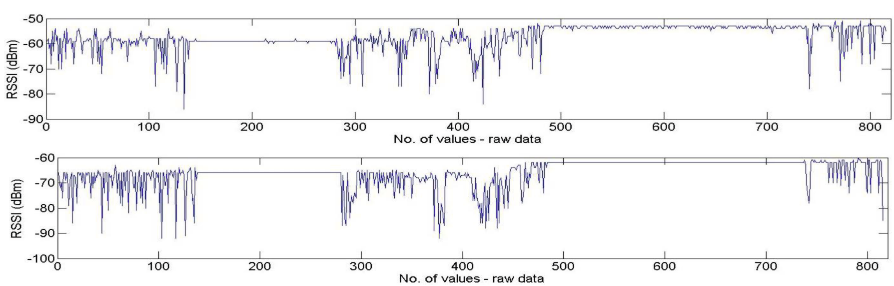



**Fig. 2** Raw data from two selected links

Figure 4 shows the threshold selection considering the normalised data. The value used in this case was $\pm 2$. Any other value above or below this threshold is considered an event which will be classified as motion. The threshold is dependent on the environment. In very noisy environments we need to modify this threshold. Thus a calibration depending on the level of noise in the environment is required.

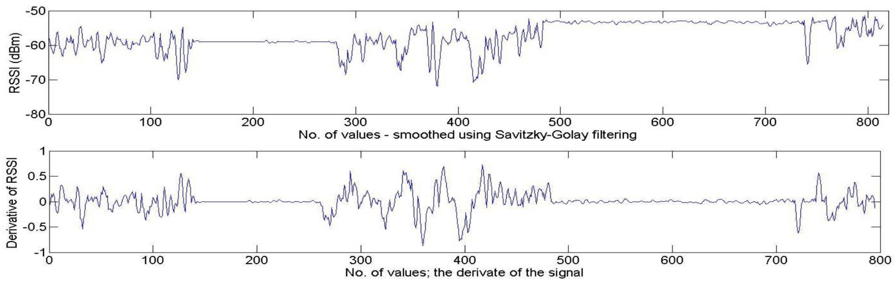



**Fig. 3** Smoothing and derivative of one link

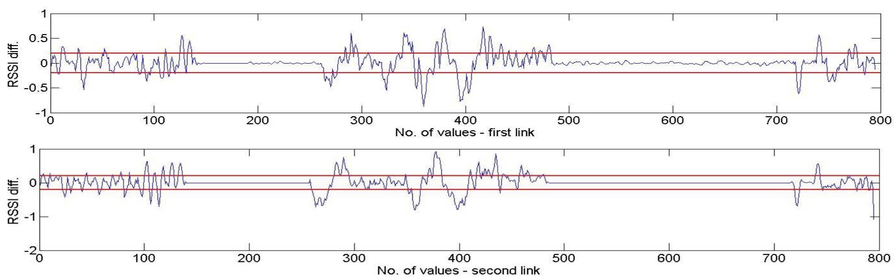



**Fig. 4** Threshold selection on the derivative of the data

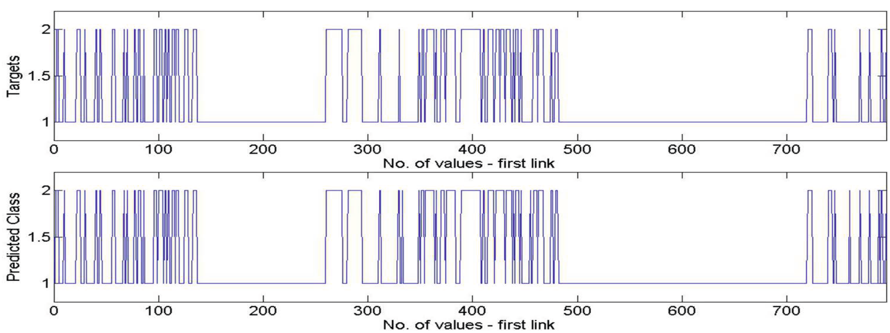



**Fig. 5** Targets and predicted classes using SVM with Polynomial kernel function

Figure 5 shows targets vector and predicted classes using SVM with the Polynomial kernel function. One link is used to train the classifiers while the data recorded on the second link represents the test vector. The targets vector is obtained by analysing the data based on the threshold chosen above. Afterwards the test data is fed to the classifier and the output is compared with the targets vector. Due to the limited space available, targets and predicted class for other kernel functions will not be added. Figure 5 represents an

**Table 1** Classification errors

| SVM Kernel Function | Error |
|---|---|
| Linear | 0.1635 |
| Quadratic | 0.0226 |
| **Polynomial** | **0.0201** |
| RBF | 0.0252 |
| MLP | 0.1686 |

example of SVM classification. As one may notice the data is classified into two classes: 'No motion' (value 1) and 'Motion' (value 2).

Table 1 shows the errors obtained in the classification process. Considering the number of values we have used for training, we can conclude that the SVM classifier performed well.

The Polynomial kernel function performed better for RSSI measurements compared to other functions. The Mean Square Error (MSE) was used to calculate the error between targets and predicted classes. The MSE obtained for SVM with Polynomial function was 0.0201.

## 4 Conclusion

In this paper we presented Support Vector Machine classifier that enables motion detection in a DfPL scenario. Various kernel functions such as Linear, Quadratic, Polynomial, Gaussian Radial Basis Function and Multilayer Perceptron were used to process wireless signal strengths in order to detect motion. The results showed the possibility of using classifiers in order to detect multi-occupancy using DfPL considering the timestamps and links affected by human presence as parameters. The usage of timestamps and links in order to decide upon the number of people in the monitored environment is considered as future work. A person cannot affect wireless links covering different areas in the environment at the same time. We analysed a bidirectional communication between two nodes in the deployed WSN. Further, more complex classifiers will be analysed in order to obtain a high accuracy motion detection.

**Acknowledgement.** This work is supported by a Vice-Chancellor's Research Scholarship from the University of Ulster. We would like to thank Oracle for their help in supplying Java Sunspot Development Kits. It is also a pleasure to thank those who helped and supported me. I am grateful to my family for their support. I would particularly like to thank my friends Marian Baciu and Sorin Vreme for their friendship and guidance. This paper is dedicated in loving memory of my father and my friend Iulian Stoi.

## References


1. Frazier, L.M.: Surveillance through walls and other opaque materials. IEEE Aerospace and Electronic Systems Magazine 11(10), 6–9 (1996)
2. Ma, L., Zhang, Z., Tan, X.: A novel through-wall imaging method using ultra wideband pulse system. In: International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2006, pp. 147–150 (December 2006)
3. Wilson, J., Patwari, N.: See-through walls: Motion tracking using variance-based radio tomography networks. IEEE Transactions on Mobile Computing 10(5), 612–621 (2011)
4. Aryanfar, F., Sarabandi, K.: Through wall imaging at microwave frequencies using space-time focusing. In: IEEE Antennas and Propagation Society International Symposium, vol. 3, pp. 3063–3066 (June 2004)
5. Gazit, E.: Improved design of the vivaldi antenna. IEEE Proceedings H, Microwaves, Antennas and Propagation 135(2), 89–92 (1988)
6. Valtonen, M., Maentausta, J., Vanhala, J.: Tiletrack: Capacitive human tracking using floor tiles. In: IEEE International Conference on Pervasive Computing and Communications, PerCom 2009, pp. 1–10 (March 2009)
7. Krumm, J.: Ubiquitous Computing Fundamentals. CRC Press (2010)
8. Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S.: Multi-camera multi-person tracking for easy living. In: Proceedings of Third IEEE International Workshop on Visual Surveillance, pp. 3–10 (2000)
9. Microsoft Research. Easy Living (2011), http://www.research.microsoft.com/
10. Deak, G., Curran, K., Condell, J.: Filters for RSSI-based measurements in a Device-free Passive Localisation Scenario. International Journal on Image Processing & Communications 15(1), 23–34 (2011)
11. Deak, G., Curran, K., Condell, J.: History Aware Device-free Passive (DfP) Localisation. International Journal on Image Processing & Communications 16(3-4), 21–30 (2012)
12. Moussa, M., Youssef, M.: Smart cevices for smart environments: Device-free passive detection in real environments. In: IEEE International Conference on Pervasive Computing and Communications, PerCom 2009, pp. 1–6 (2009)
13. Youssef, M., Mah, M., Agrawala, A.: Challenges: device-free passive localization for wireless environments. In: Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking, pp. 222–229 (2007)
14. Kosba, A., Abdelkader, E., Youssef, A., Analysis, M.: of a device-free passive tracking system in typical wireless environments. In: 2009 3rd International Conference on New Technologies, Mobility and Security, NTMS, pp. 1–5 (December 2009)
15. Song, L.-P., Yu, C., Liu, Q.H.: Through-wall imaging (twi) by radar: 2-d tomographic results and analyses. IEEE Transactions on Geoscience and Remote Sensing 43(12), 2793–2798 (2005)
16. Mathworks. R2012a Documentation, Bioinformatics Toolbox (2012), http://www.mathworks.co.uk/help/toolbox/bioinfo/

# Obtaining Internet Flow Statistics by Volunteer-Based System

Jens Myrup Pedersen and Tomasz Bujlow

Department of Electronic Systems, Aalborg University, Denmark
{jens,tbu}@es.aau.dk

**Summary.** In this paper we demonstrate how the Volunteer Based System for Research on the Internet, developed at Aalborg University, can be used for creating statistics of Internet usage. Since the data is collected on individual machines, the statistics can be made on the basis of both individual users and groups of users, and as such be useful also for segmentation of users intro groups. We present results with data collected from real users over several months; in particular we demonstrate how the system can be used for studying flow characteristics - the amount of TCP and UDP flows, average flow lengths, and average flow durations. The paper is concluded with a discussion on what further statistics can be made, and the further development of the system.

## 1 Introduction

Understanding the behavior of Internet traffic is crucial in order to model traffic correctly, and to create realistic scenarios of future Internet usage. In particular, understanding the behavior of different kinds of traffic makes it possible to create scenarios of increasing/decreasing particular amounts of traffic. The models can then be used for analysis and/or simulations of distribution and backbone networks under different scenarios. The application of different provisioning and traffic engineering techniques can be tested as well.

Traffic statistics are today often made by Internet Service Providers (ISPs), who monitor the activity in their networks. However, often ISPs consider these data to be private and are not keen on sharing with researchers. Some traces are publicly available, such as the Caida data sets[1]. Even with access to traces from ISPs or other, traffic monitored in the network cores suffers from missing important statistics that can only be known accurately at the sources - such as inter-arrival times between packets and flow durations. It should be noted that the literature covers a number of interesting studies, where researchers have gained access to real data, e.g. [2] and [3]. In the latter, the

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 261–268.
springerlink.com 

data are collected at the broadband access router, which is quite close to the generating source.

There are also a large number of commercial tools for monitoring traffic in Local Area Networks (e.g. on Internet gateways), such as Cisco Netflow [4]. These can provide useful insights to traffic, but without collecting traffic from many different networks it does not give a good overview of what Internet traffic looks like.

The Volunteer Based System (VBS) for Research on the Internet, developed at Aalborg University, seeks to avoid these problems by collecting traffic from a large number of volunteers, which are agreeing to have their Internet usage monitored and statistics collected for research purposes. This provides statistics from the point where the traffic is generated, meaning that quite precise statistical data can be obtained. Moreover, it is possible to monitor also which applications are opening the sockets, and thus get a precise picture of the behavior of different applications. The general idea was first described in [5] and a preliminary limited prototype was implemented in [6]. The current system design was announced in [7], while more technical details on later refinements can be found in [8]. Other papers ([9], [10] and [11]) demonstrate various applications of our system.

In this paper we show how the system can be used for generating statistics at flow level. The paper is organized as follows: First, in Section 2 we describe how the data collection is made and how statistics are extracted. In Section 3 we present the results, and in Section 4 we conclude the paper and discuss the further work.

The authors would like to stress that the system is based on open source software, published on Source Forge [12]. We would like to take the opportunity to encourage other researchers to use the system for collection of Internet traffic information, and to the widest possible extend share the data traces with the research community.

## 2 Data Collection and Extracting Statistics

In this section we briefly describe the fundamentals of VBS, with a particular focus on the parts that influence the monitoring, data collection, and extraction of statistics. For more details, please refer to our previous paper [8].

For each volunteer, the system monitors both ingoing and outgoing traffic on his/her computer. Storing all this data, and transferring it back to the server, would be a huge task, if not impossible given the limited upload capacity available on many standard Internet connections. Therefore, the data is saved as follows, and transmitted to our central server as:

- For each flow, information is stored about e.g. source and destination IP addresses and port numbers, flow start time, flow end time, number of packets, protocol (TCP or UDP) as well as a flow ID. Moreover, information is collected about which process has opened the socket, a feature

which provides valuable information about the characteristics of traffic created by different applications.
- For each packet the main information is the packet size and the relative time stamp. Moreover flags from the header is stored, as well as information of packet direction and flow ID.

In this way all relevant information is stored, while the requirements in terms of memory and network usage is kept at a minimum. Also, no payload is stored at any time which is an advantage with respect to privacy and security. One privacy concern has been the transfer of source and destination IP addresses. In the current implementation, the IP addresses are hashed before being transferred to the server. However, since the hash function is known (open source), and since the number of IP addresses in IPv4 is limited, it is not difficult to determine the original IP address.

The purpose of this particular study was to demonstrate the usage, so focus was on obtaining and presenting data from a limited number of users prior to run more large-scale experiments. The statistics were obtained from 4 users during the period from January to May 2012. One of the four users (User 4) did not join the system until late April, and thus only participated for the last 2-3 weeks of the study. Due to being a heavy user, the amount of data collected from this machine is higher than from any of the other participants, despite the shorter participation. During the time of study all the traffic from the users were collected by the system as described above, and the data stored into our central database.

The four users can be described as follows:

- User 1 - Private user in Denmark
- User 2 - Private user in Poland
- User 3 - Private user in Poland
- User 4 - Private user in Denmark

With the data collected, a wide variety of studies can be conducted. For this paper, we chose to analyze only the flow data (not the packet data), since the amount of data makes it more manageable. As the main purpose is to demonstrate the usefulness of the system, we chose to derive the following statistics.

- Amount of TCP and UDP flows
- Average flow lengths for TCP and UDP flows
- Average flow durations for TCP and UDP flows
- Top 5 applications (measured on the number of flows)

The statistics are done for the individual users as well as for the users altogether.

# 3 Results

## 3.1 TCP and UDP Flows

The distribution of TCP and UDP flows are shown in Table 1. It can be seen that both the number of flows and the distribution between TCP and UDP varies quite a bit between the different users.

**Table 1** The numbers of TCP and UDP flows for all users as well as for the individual users. The number in parenthesis shows the distribution.

| User | #UDP flows | #TCP flows |
|---|---|---|
| All | 4770315 (55%) | 386530 (45%) |
| 1 | 446692 (35%) | 820927 (65%) |
| 2 | 3142581 (60%) | 2084590 (40%) |
| 3 | 693389 (52%) | 642740 (48%) |
| 4 | 487653 (60%) | 315273 (40%) |

## 3.2 Flow Lengths and Durations

The distribution of flow lengths (number of packets per flow) for TCP and UDP for the different users are shown in Table 2. It is quite interesting to observe that the flow lengths for both TCP and UDP are so different between the different users, indicating a different Internet usage. It should be noted that with the data in the system it is possible to make a more detailed analysis of the distribution of flow lengths, not only for the different users but also for each application used by each user.

**Table 2** Average flow lengths for TCP and UDP flows

| User | Avg. UDP length | Avg. TCP length |
|---|---|---|
| All | 72 | 81 |
| 1 | 5 | 110 |
| 2 | 90 | 80 |
| 3 | 34 | 29 |
| 4 | 72 | 114 |

The distribution of flow durations (in seconds) is shown in Table 3. It seems that for users 1-3 the users who have longer average flows also have longer average flow durations. However, user 4 seems to have quite short flow durations even though the flows are quite long. Even though a more thorough

**Table 3** Average flow durations for TCP and UDP flows

| User | Avg. UDP duration | Avg. TCP duration |
|---|---|---|
| All | 33 | 26 |
| 1 | 1 | 32 |
| 2 | 41 | 25 |
| 3 | 28 | 7 |
| 4 | 19 | 55 |

analysis is required to explain this in detail, we assume it is due to the user being on a fast Internet connection. However, the type of traffic with generally longer flows probably also plays a role.

### 3.3 Top 5 Applications

Analyzing the applications is more challenging than deriving the other parameters. First, we did not manage to collect the socket name for a substantial amount of the flows. This is mainly concerning very short flows, where the opening time of the socket is so short that it is not captured by the socket monitor. Secondly, what we obtain in order to identify an application is really the process name. For this study 240 different process names were identified. Further work is needed in order to group these into applications, and for this study we just list the top 5 process names. It should be noted that it is not a trivial task to determine how e.g. browser plugins should be grouped and categorized.

The top application names for the different users are shown in Tables 4-8.

Based on the information obtained by the system it is possible to make additional statistics, taking e.g. the flow lengths of different applications into account. Also packet statistics (e.g. packet lengths) can be taken into account, providing a quite precise picture of what applications are taking up most bandwidth for the different users.

Without going into a more detailed data analysis, we did an observation regarding the unknown flows, which is worth highlighting. The Unknown flows

**Table 4** Top 5 applications for all users

| Application name | Number of flows | % of all flows |
|---|---|---|
| uTorrent | 6399336 | 74.12 |
| Unknown | 948497 | 10.99 |
| Chrome | 441953 | 5.12 |
| Firefox | 361213 | 4.18 |
| Svchost | 103757 | 1.2 |

**Table 5** Top 5 applications for user 1

| Application name | Number of flows | % of all flows |
|---|---|---|
| Unknown | 729868 | 57.56 |
| Firefox | 330498 | 26.07 |
| Chrome | 138105 | 10.89 |
| Amule | 18863 | 1.49 |
| Ntpd | 17929 | 1.41 |

**Table 6** Top 5 applications for user 2

| Application name | Number of flows | % of all flows |
|---|---|---|
| uTorrent | 4674545 | 89.43 |
| Chrome | 227616 | 4.35 |
| Unknown | 136387 | 2.61 |
| Svchost | 101633 | 1.94 |
| Java | 38704 | 0.74 |

**Table 7** Top 5 applications for user 3

| Application name | Number of flows | % of all flows |
|---|---|---|
| uTorrent | 1220728 | 91.36 |
| Chrome | 76062 | 5.69 |
| Unknown | 21764 | 1.63 |
| SoftonicDownloader | 15035 | 1.13 |
| Java | 2169 | 0.16 |

**Table 8** Top 5 applications for user 4

| Application name | Number of flows | % of all flows |
|---|---|---|
| uTorrent | 504063 | 62.78 |
| Moc | 90358 | 11.25 |
| Iexplore | 64407 | 8.02 |
| Unknown | 60478 | 7.53 |
| Firefox | 26896 | 3.35 |

account for a large amount of the total flows. However, the flows have an average length of 2 seconds and an average of 11 packets, indicating that it is not such a large share of the total traffic. These unknown flows are almost equally shared between TCP(53%) and UDP(47%).

### 3.4 Cumulative Number of Flows

Distribution of cumulative number of flows for the 4 users during the time of our experiment is shown in Fig. 1.

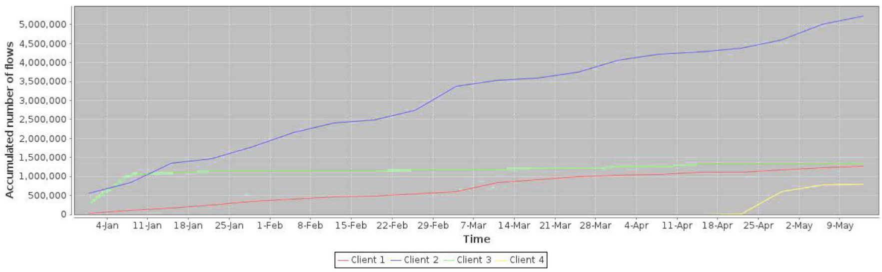



**Fig. 1** Cumulative number of flows for all users

## 4 Conclusion and Discussion

In this paper we have demonstrated how the Volunteer Based System for Research on the Internet developed at Aalborg University can be used for creating statistics of Internet traffic, specifically within the studies of flows and their properties.

Future research will focus on developing efficient methods for extracting relevant information from the packet statistics. This can provide even more valuable information about the flows, for example on average packet sizes of different flows (and distribution of packet sizes), inter-arrival times between packets, and the number of successful vs. unsuccessful connections for different kinds of traffic. Moreover, particularly interesting statistics can be derived from the combined flow and packet statistics, such as the average size of flows of different kinds of traffic, and eventually how much traffic is created by different applications for individual users. The challenge is that it is large amounts of data, so efficient ways of handling these has to be developed.

Another important part is recruitment of more volunteers, in order to collect larger amounts of data. Also, having appropriate background information about the users could be useful. This includes both data about the use themselves such as age, occupation, if the computer is shared etc., but also information about the connection, e.g. speeds and technologies.

In order to obtain more data, other researchers are invited to join the project and use it for collection of data for scientific purposes. The code is available as open-source, and can be found together with a comprehensive documentation on our project homepage [12] located on SourceForge.

## References

1. http://www.caida.org/data/passive/
2. Fraleigh, C., Moon, S., Lyles, B., Cotton, C., Khan, M., Moll, D., Rockell, R., Seely, T., Diot, C.: Packet-Level Traffic Measurements from the Sprint IP Backbone. IEEE Network 17, 6–16 (2003)
3. Maier, G., Feldmann, A., Paxson, V., Allman, M.: On Dominant Characteristics of Residential Broadband Internet Traffic. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, pp. 90–102 (2009)
4. http://www.cisco.com/go/netflow
5. Balachandran, K., Broberg, J.H., Revsbech, K., Pedersen, J.M.: Volunteer-Based distributed traffic data collection system. In: Proceedings of the 12th International Conference on Advanced Communication Technology, pp. 1147–1152 (February 2010)
6. Balachandran, K., Broberg, J.H.: Volunteer-Based Distributed Traffic Data Collection System. Master Thesis at Aalborg University, Department of Electronic Systems (June 2010)
7. Bujlow, T., Balachandran, K., Riaz, T., Pedersen, J.M.: Volunteer-Based System for classification of traffic in computer networks. In: IEEE 2011 Proceedings of the 19th Telecommunications Forum TELFOR 2011, pp. 210–213 (November 2011)
8. Bujlow, T., Balachandran, K., Hald, S.L., Riaz, T., Pedersen, J.M.: Volunteer-Based System for research on the Internet traffic. TELFOR Journal (accepted, 2012)
9. Bujlow, T., Riaz, T., Pedersen, J.M.: A Method for Assessing Quality of Service in Broadband Networks. In: Proceedings of the 14th International Conference on Advanced Communication Technology, pp. 826–831 (February 2012)
10. Bujlow, T., Riaz, T., Pedersen, J.M.: A Method for classification of network traffic based on C5.0 Machine Learning Algorithm. In: Proceedings of ICNC 2012 International Conference on Computing, Networking and Communications (ICNC): Workshop on Computing, Networking and Communications, pp. 244–248 (February 2012)
11. Bujlow, T., Riaz, T., Pedersen, J.M.: Classification of HTTP traffic based on C5.0 Machine Learning Algorithm. To appear in Fourth IEEE International Workshop on Performance Evaluation of Communications in Distributed Systems and Web-based Service Architectures, PEDISWESA 2012 (2012)
12. http://sourceforge.net/projects/vbsi/

# A Concept of Unobtrusive Method for Complementary Emotive User Profiling and Personalization for IPTV Platforms

Adam Flizikowski[1], Mateusz Majewski[2], Damian Puchalski[2], Moustafa Hassnaa[3], and Michał Choraś[1]

[1] University of Technology and Life Science, Kaliskiego 7, Bydgoszcz, 85-796, Poland
adamfli@utp.edu.pl

[2] ITTI, Rubiez 46, Pozna, 61-612., Poland
mateusz.majewski@itti.com.pl, damian.puchalski@itti.com.pl

[3] France Telecom - Orange, 3840 rue General Leclerc 92794, Issy les Moulineaux, Cedex 9, France
hassnaa.moustafa@orange-ftgroup.com

**Summary.** Mobile technologies, new interactive applications and the service providers' customer-centric approach are influencing the way of assessing QoE nowadays. Traditional QoE assessment methods proved to be effective when dealing with legacy audio/video services; however, current IPTV services provide features beyond traditional TV and are not limited to delivering audiovisual content but may also rely on auxiliary services (e.g. content recommendation). Personalization mechanisms that learn instantaneous user-context relation are interesting extension of the QoE parameters enabling improved experience customization. This paper is focused on the QoE-context relation for context-aware IPTV platforms offering personalized TV experience. The latter systems are in the scope of the UP-TO-US[1] project which is treated in this paper as a reference project dealing with user experience and IPTV. Authors define a QoE architecture for validating traditional subjective assessment methodologies (e.g. based on human visual system modeling, or standardized methodologies like ITU-T BT.500-11) by adopting additional context characteristics - user emotions. Moreover the proposed QoE module is aligned with the architecture defined in the UP-TO-US. In the proposed approach to affective QoE authors foresee important role for learning algorithms that can be applied in order to build a user model (an agent reasoning on QoE based on the gathered knowledge about user-content relation).

## 1 Introduction

Quality of Experience (QoE) is defined by the International Telecommunication Union (ITU) as: "the overall acceptability of an application or service,

[1] UP-TO-US is a European project under the Eureka Celtic Programme.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 269–281.
springerlink.com 

as perceived subjectively by the end-user" [1]. Traditional methods for assessing QoE include subjective [2][3] and objective [4] methods, both take into account only user's audio/visual perception perspective. ITU also provides an extension to the QoE notion, stating that the term Quality of Experience "includes the complete end-to-end system effects (client, terminal, network, service infrastructure, etc.)" and that the "overall acceptability may be influenced by user experience and context"[1]. There are many factors which have a direct influence on service quality as perceived by end-user (including human audio/visual impairments). It seems that improved QoE assessment shall derive from the user-context relation. The affective state of the human is an important, direct source of information and feedback that influences human behavior and social interactions [5]. Emotions govern the perception of people, what they think of an object/content/service and how they decide to act/behave in response to a stimulus.

In turn the following can be stated: emotions drive user preferences and choice. In particular the purely reactive affective state of the human brain influences the selection of factors that are of key relevance for individuals ("this item satisfies my requirements") and which are less important or even irrelevant ("this item does not match my taste"). Thus it is important to analyze user emotions and use the results of this analysis to develop models that mimic user experience [6] and provide additional information for recommendation systems. Such information based on context information, such as movie metadata and (traces of) emotions experienced by a user, can increase the accuracy of recommender system [7]. Improved recommendations in turn has influence on user's experience while interacting with the system (as the content offered to the user is carefully selected based on his affective states).

Authors' hypothesis is that there exists a relation between the user psychological reaction and the audiovisual content that affects the viewer score of the content - so that based on the nature of the relation the user experience is also affected. Thus analysis and modeling of such relation can be used to improve legacy recommender systems. In particular the content being viewed changes the user's affective state (it makes the user happy/sad, excited/frustrated, etc.). Thus assessing the emotion-content relation and applying learning algorithms to recognize the preferences of the user could lead to better service customization (e.g. "this types of scenes satisfy the user requirements"). In particular the emotional response of the user to different scenes could be registered and stored to build a knowledge database about the user profile.

Authors plan to prove the above hypothesis by performing the following activities:

I) Comparing existing results of subjective tests based on human perception model (system-oriented) to test with affective analysis (user-centric) - in particular we specify a method for collecting a subject's brainwave data during subjective video quality assessment to learn the content-emotion correlation function and to learn how particular content influences user quality ratings;

II) Tagging audiovisual content to be tested with extended metadata (e.g. description of scenes and emotions they typically evoke);

III) Proposing an automated user's personality identification mechanism using external EEG (Electroencephalography) reading device. The results of such identification would help solving the typical "cold start problem" of a recommender systems that causes mismatching recommendation if the user is not accurately identified;

IV) The model for assessing QoE for audiovisual content should take the user's context into account - in particular we plan to show how user's context (e.g. his lifestyle, behavior) influences quality perception experienced by the user.

This paper defines framework that will eventually enable means of proving the above hypothesis in the future research. The paper is structured as follows: Section 2 provides state of the art research on topics related to QoE assessment and recommender systems, Section 3 introduces the goals of UP-TO-US project; Section 4 describes the method of affective user assessment; the architecture for the QoE module is provided in Section 5; we conclude with Section 6.

## 2 Related Work

The author in [8] proposes a QoE model that consists of three major components: sensorial evaluation, perceptual evaluation and evaluation of affective state of a person. According to this paper, these three perspectives are crucial in the process of video service evaluation and determination of a user's general satisfaction.

Legacy approach to perceptual evaluation uses subjective or objective assessment methods. Subjective tests provide the research community with valuable information on how users perceive video/voice quality and have to follow ITU-T recommendations to produce most reliable (and comparable) results. An interesting elaboration of common mistakes made by researchers during subjective assessment methodology preparation can be found in [9]. On the other hand the legacy subjective methodologies do not take the user context into account. For example - there is a substantial difference between a test subject viewing a movie at home and in a controlled lab environment. An interesting approach to subjective assessment of QoE can be found in [10] - test participants were given a full length movie on DVD to be viewed in their home environment; the movie had several scenes inserted with degraded quality and the test subjects were not aware of the existence or timing of these scenes; during a traditional subjective test users were told to rate the video quality, which implies that they should focus on the quality of particular scenes. The latter however stimulates the participants to concentrate on the video quality rather than the content. In contrast the use of full length movies is a more realistic approach (context-aware) that may cause subjects not to notice an impaired scene during movie consumption (because they might be emotionally affected by the movie plot and thus not pay so much attention to the video artifacts) - in turn no influence on user's perceived QoE should be observed. Consequently, authors in [10] suggest that user expectations and context both have influence on QoE measurement based on perceived video quality. Similar observations were highlighted in [11], where authors tried to assess the influence of sensory effects on perceived quality. The results show that users tend to rate the quality of same video samples higher when sensory effects are present. That is why some authors focus on the evaluation of user emotions (emotions might help traceinga brain's answer to the sensory effects). Many projects and experiments prove that analysis of human emotions is complementary to legacy evaluations in the area of services' quality assessment, particularly in the entertainment services domain (i.e. for pictures, clips, movies and music quality of experience assessment). Such analysis can be used for automated tagging of content in order to enhance capabilities of content recommendation systems based on individual and personalized users' preferences [7]. In order to be successful with the above mentioned affective analysis an efficient methods for classifying user emotions in an unobtrusive way are needed. In the literature several methods for capturing emotions online is mentioned. These methods are based on galvanic skin response (GSR), face recognition and mapping user brain waves into emotions. The authors in [12] propose tracking of emotional states (based on face recognition) together with assessment of attention, perception, interaction or skills level in order to implement QoE adaptation in the e-learning systems. In [7] the author proposes utilizing user's emotions to build a content-based recommender for multimedia resources which are described by affective (emotional) metadata acquired in a non-invasive manner using face recognition through video camera. This work has shown that awareness of user emotional response to a given stimuli, together with appropriately labeled content, can increase the accuracy of the content recommendation system. In particular the emotional traces of a user can be used to solve the "cold start problem" of the CBF-based (collaborative filtering) recommender algorithm. However the approach mentioned above targets recommending still images only. Human mental states which are detected by analyzing the subjects' face expression recordings are also considered in [13]. Research documented in [14] delivers results of experiment, in which face expression detection and automated analysis of body language are used for emotion recognition. The emotions detected are induced by video clips containing specific emotional load (e.g. causing anger, sadness, happiness, joy or neutral).

Contrary to the above mentioned work the authors in [15] note the fact that popular methods for emotions detection based on face recognition techniques may be insufficient, because face expressions are relatively easy to control, regardless of perception and emotions. For more reliable and adequate results of emotion assessment, signals which are more spontaneous and less vulnerable to intentional control should be considered. Examples of such signals are those derived from peripheral nervous system (e.g. heartbeat rate changes, EMG - electromyography, GSR- Galvanic Skin Resistance, etc.). In [8] the author highlights that emotion may be characterized both by positive and negative features; however, such simplification may lead to inappropriate results of content analysis due to the different target emotional stimulation depending on the movie type (e.g. fear which is a negative emotion is a desired brain response to a horror movie). Author indicates that target preference analysis system should take into account the type of video content and the level of emotions which are adequate for this particular type of content and which are adequate to particular user's preferences/profile. The author suggests that the heartbeat rate, muscles activity, skin conductance and temperature are used to measure user emotional response to content. However author in [8] does not define specific metrics for affective QoE assessment - only general equations with specific weights for sensorial, perceptual and emotional dimensions are included.

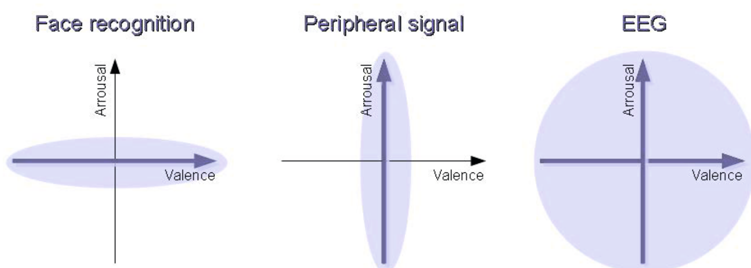



**Fig. 1** Three dimensions of emotions measurement in the two-dimensional space

The previously mentioned techniques for emotion recognition provide only one-dimensional emotional information (valence or arousal) on a two-dimensional emotional space model proposed in [16]. Mapping of the EEG signals into emotions, however, provides information on the type of emotion (valence axis) and the intensity of the emotion (arousal axis). The comparison between different methods is depicted in Fig. 1.

Authors in [15] describe an experiment in which methods for collecting non-voluntary responses of the human body, like EEG, GSR, skin temperature and blood pressure, are used to recognize viewer's emotions when being presented with pictures from the IAPS database [17]. After collecting data from a EEG device, participants performed a self-assessment and evaluation of emotional valence and arousal level based on numerical scale (with five possible numerical judgments) in order to validate results from EEG.

The authors in [18] also propose emotions detection based on EEG analysis. In the described experiment, EEG signal was a starting point for further detection of eye blinking frequency and intensity during playing a racing video game. The purpose of the experiment was to prove that EEG provides capabilities for determination of logical interdependencies between EEG signals and level of concentration during video games.

The above analysis proves that user emotions are an important factor (complementary to legacy ones) when considering audiovisual services. In this light, full length movies can be viewed as a dynamic system which evokes emotions. Authors of this paper are aware of the gap in current research in the area of effective methods for assessing user's emotions during audiovisual content consumption. The above mentioned related work can be concluded by saying that an appropriate tagging mechanism that is grouping movie scenes based on user emotion data is an interesting direction towards more comprehensive modeling of QoE.

However such an approach requires the implementation of an appropriate content tagging mechanism for full length movies. The concept of content tagging with extended metadata describing particular movie scenes is presented in Fig. 2.

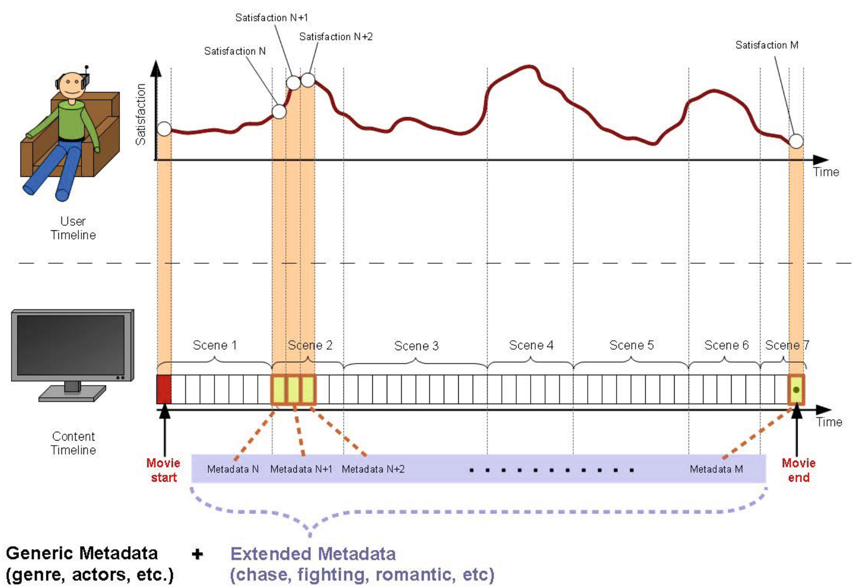



**Fig. 2** Movie with extended metadata

The content metadata and the captured user emotions could be then aligned in time and fed into a learning algorithm to learn user preferences based on the emotion-scene relation.

## 3 UP-TO-US Project Description

The advances in IPTV technology pave the way for improved service provisioning model, moving from traditional broadcastercentric TV services to a new usercentric TV model. Such a new model will allow users not only to access new services and functionalities from their providers, based upon their profiles and context of service consumption but also to become active part in the content personalization through contributing to building their dynamic profiles. This IPTV model is promising in allowing low cost services for endusers and a revenue system for broadcasters based on personalized advertising methods, as well as new business opportunities for network operators and service providers. The objective of the UPTOUS (User-centric Personalized IPTV ubiquitous and secure services) project [19] is to elaborate, prototype, and evaluate an open European solution allowing IPTV services personalization over different IPTV systems (having different architectures and belonging to different network operators and service providers), through content personalization according to a user context and the environment's context (network and devices) while preserving viewer privacy. UPTOUS focuses on two usecases for service personalization: users in nomadic situations in a hotel, in a friend's home or anywhere outside a domestic sphere (allowing the user to access his personalized IPTV content in a hotel for instance and be billed on his own bill "My Personal Content Moves with Me"), and users' mobility in his domestic sphere (allowing the user to move around within his domestic sphere while continuing accessing his IPTV service personalized according to his location and devices in his proximity ("My Content Follows Me in a Customized Manner"). In order to achieve the objective of UPTOUS, some enabler technologies will be developed and integrated to different IPTV systems (that will be chosen during the project). These mainly include: i) a contextaware module capable of monitoring and gathering the user and his environment contexts and feeding them in a dynamic manner to the IPTV system, ii) a QoE management module tackling the users' semantic level (emotion and its influence with the user context) and the resource level (network and devices QoE related factors), iii) a profiling management module, capable of constructing and dynamically updating the users profiles according to the various contexts, and iv) a privacy management module that will be responsible for managing the different privacy levels for each user and protecting the user personal information. Consequently, personalized services could be provided in which content is selected according to users' preference, Quality of Experience (QoE) requirements, and different contexts, while fostering trust between viewers and broadcasters through an efficient privacy management, and thus encouraging viewers to participate actively in this interactive usercentric TV paradigm through allowing their continuous contexts gathering.

## 4 Subjective User Assessment

The authors of this paper are planning as a future work to prepare and carry out subjective assessment tests with end users to support the ideas of emotional profiling system presented in the section above. In particular, most promising existing subjective assessment methodologies for audiovisual content [2] will be implemented and coupled with the user affective assessment layer. At the preliminary stage the proposed methodology includes surveying end-users through various questionnaires and organized focus groups with the users. The goal of this activity is to gather preliminary information on end user's requirements, constraints and on the correlation of emotional states and content being watched. Once such data is available, subjective assessment can provide positive feedback to calibrating the process of capturing human brain activity in a measurement campaign with real users. Approach undertaken in this paper is based on the assumption that a positive correlation between the content and the IPTV service user exists. To prove the latter a human-content relation has to be evaluated by repeating numerous subjective assessment runs with end-users. The following general steps (depicted in Fig. 3) are planned to be performed in order to capture data for analysis and evaluation of the role of emotions in content delivery systems:

1. Research on related work will be conducted,
2. Suitable test methodology will be prepared,
3. Surveying users - delivering questionnaires to test subjects and collecting the results (an example of such a questionnaire would be the user's temperament questionnaire to differentiate the users [20]),
4. Video content and lab environment preparation - content will be prepared according to the preferences collected in point 3. A video sequence will be assembled to match different tastes of test subjects (as presented in the lower part of Fig. 2),
5. Test subjects from point 3 will be invited to conduct subjective assessment in a controlled environment (lab),
6. The main part of the test will be performed according to subjective assessment block in Fig. 3,
7. The video sequence from point 4 will be presented to each test subject and EEG devices will be used to intercept signals induced by the test subject's brain,
8. The resulting database of brain signal traces will be mapped into emotional state (as presented in the upper part of Fig. 2),
9. After subjective assessment users will fill out questionnaires to validate the results and test methodology (this will be used to redefine the test methodology),

10. A correlation between content, user and emotions will be analyzed,
11. Points 2 to 10 will be repeated (on a regular basis) to validate the results.

After subjective tests are completed a group of representative test subjects will be selected for participating in the evaluation process for additional tests. We assume that some of the users under test will be involved in subjective assessment in their home environment to evaluate consistency of results obtained in the lab environment (as specified in the lower part of Fig. 3). It might also be interesting to perform the tests with different groups of people, especially in different geographical regions, taking into account their cultural heritage, religion and place of origin. It will be a challenging task to assess human emotions as there are many factors that could have direct influence on the affective state of the user (degree of immersion, mental workload, stress, etc.). Therefore, it is required to perform user-EEG device calibration (either in form of yes/no' questions or software calibration) before each individual subjective test. These results will be used as a reference point for future performance validation and individual user-device calibration. Concluding, the authors are going to verify if it is possible for IPTV services QoE assessment to collect information on users' preferences based on the user affective state (emotions) without the need of direct feedback from the user. Thus the next section will materialize the analysis done so far by introducing the architecture design of a QoE module that would enable processing of legacy QoE complemented with envisaged affective enhacement.

## 5 QoE Module Architecture

The result that will be obtained during subjective assessment will be used in development of the Quality of Experience (QoE) management module for affective user experience analysis within target audiovisual system. Results collection can be achieved by gathering data from various enablers in the network domain and sensors in the user domain in order to calculate and deliver QoE metrics necessary to evaluate the service quality. With the respect to the UP-TO-US context we also propose to enhance standard QoE objective assessment methods (audiovisual quality, usability, accessibility, etc.) with user emotional data.

Additionally, learning methods will be identified and applied to automate the process of constructing a (user) model for improved QoE prediction and assessment for IPTV through a learning phase that could be considered as a part of the system configuration. The results from aforementioned objective and subjective measurements extended with human component analysis will validate the role of human factor in the IPTV QoE holistic view. A detailed architecture of the QoE module is provided in Fig. 4. The QoE module is designed to provide ratings (quality, content, usability, etc.) related to an instantaneous or aggregated end-user experience and service being used. In

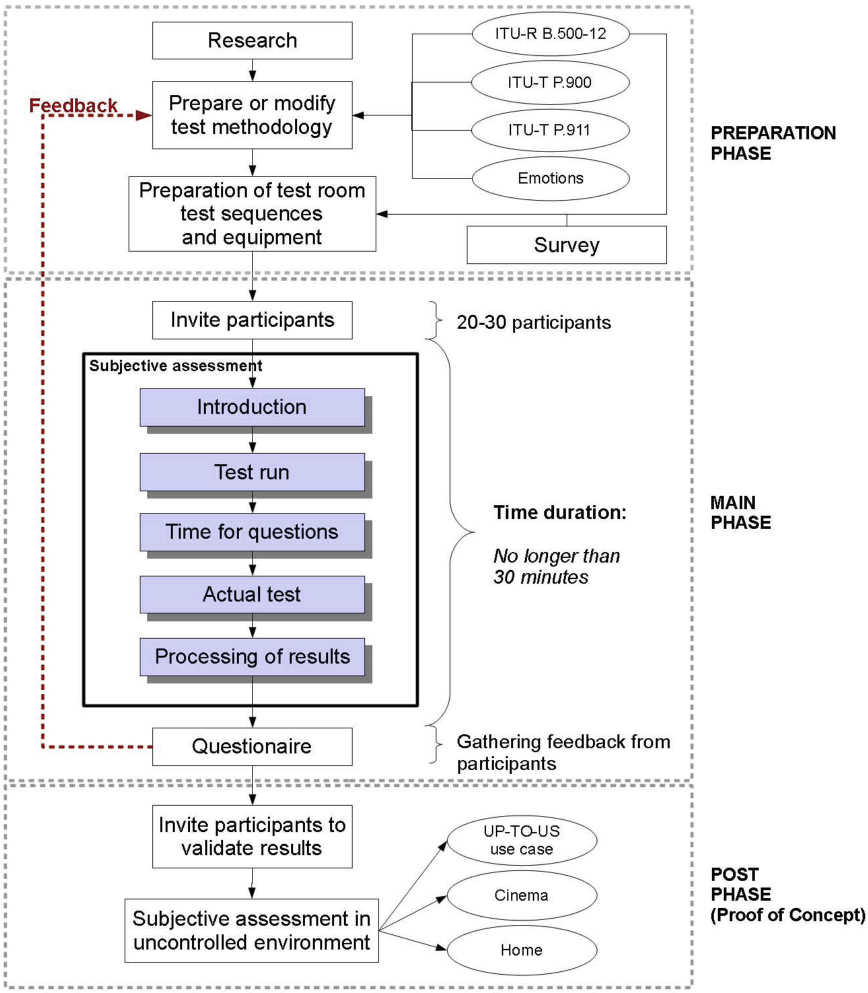



**Fig. 3** Subjective assessment methodology

particular our research will be focusing solely on services dealing with audio-visual content (LIVE TV, VoD "Video on Demand"), where the QoE module should be provided with content metadata from a context-awareness system, in order to extract information on user preferences, considering also the emotional information detected.

The module should receive input from the Context-Awareness System, User Profile Manager and external EEG-reading device. In the initial phase of development the EEG-reading device can be considered as an external data source directly connected to the QoE module (for simplicity of data collection), but in the future implementation the brainwave data should be directly fed to the Context-Awareness System. The Context-Awareness system should at least provide the QoE module with the information on the underlying network QoS (related to the user-service tuple), device type and content metadata (as indicated by the input interface in Fig. 4). Additional service-specific and user-specific data will also be provided as input to the QoE module. The output of the QoE module block is a range of different quality rating metrics. The metrics are derived from objective measurements, user QoE preferences, and user profiling (based on IAPS dataset [7] and EEG readings). In addition the module should output the emotional response of a given user to the particular content (aligned in time with the content metadata) and user preferences on selected content.

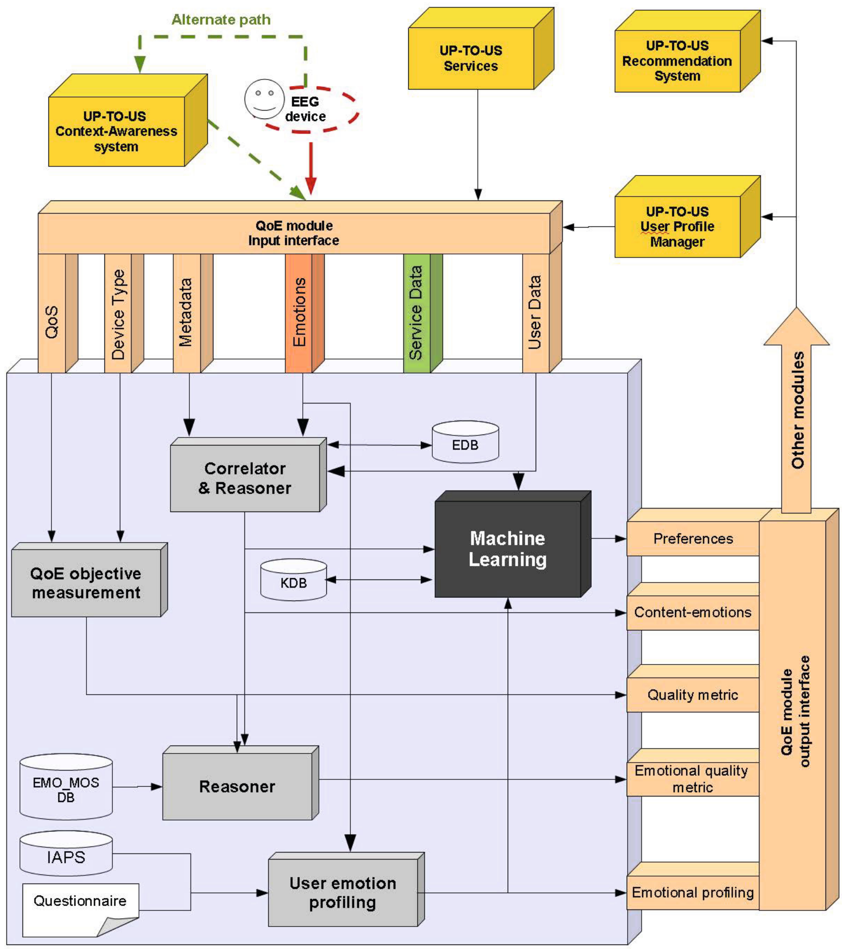



**Fig. 4** QoE module architecture

To achieve this, the target movie scenes will be tagged with extended metadata (as indicated in Fig. 2). The QoE module interacts with the UP-TO-US Recommendation System and UP-TO-US User Profile Manager. The UP-TO-US User Profile Manager will be enriched with user personality information. This information could be used by the UP-TO-US Recommendation System to solve the 'cold start problem' of CBF-based algorithms [7].

## 6 Conclusions

In this paper we presented a state of the art solutions on QoE for multimedia and audiovisual services and outlined the architecture and research activities planned for the effective QoE module that can be integrated into multimedia systems like the one defined by European project UP-TO-US. In particular authors proposed a new user QoE assessment mechanism for audiovisual content that takes user's emotions into account. In our belief important user QoE related data can be extracted and analyzed with the support of machine learning algorithms that will learn user's emotion-content relation, where the user's emotion is acquired in real-time by the context-awareness system similar to the one defined within UP-TO-US. The user QoE learning algorithms are foreseen to be part of the QoE module.

**Acknowledgement.** The work done towards preparation of this paper benefits from the cooperation with the UPTOUS European project under the Eureka Celtic Programme during the period Q4 2010 - Q2 2011.

## References

1. Recommendation ITU-T P.10 G.100: Amd.1 (2007), New Appendix I Definition of Quality of Experience (QoE) (2006)
2. Recommendation ITU-R BT.500-12: Methodology for the subjective assessment of the quality of television pictures
3. Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications (2008)
4. Furth, B.: Handbook of Multimedia for Digital Entertainment and Arts. Springer (2009)
5. Google Tech Talks (2008), `http://www.youtube.com/watch?v=tShDYA3NFVslist=QLfeature=BF` (last visited: March 25, 2011)
6. Tsianos, N., Germanakos, P., Lekkas, Z., Mourlas, C.: Evaluating the significance of cognitive and emotional parameters in e-learning adaptive environments. In: IADIS International Conference on Cognition and Exploratory Learning in Digital Age (2007)
7. Tkalcic, M.: Recognition and usage of emotive parameters in recommender systems. PhD Thesis, University of Ljubljana (2010)

8. Pereira, F.: A Triple User Characterization Model for Video Adaptation and Quality of Experience Evaluation
9. Keimel, C., Oelbaum, T., Diepold, K.: Improving the verification process of video quality metrics. In: International Workshop on Quality of Multimedia Experience, QoMEx 2009, pp. 121–126 (2009)
10. Staelens, N., Moens, S., Van den Broeck, W., Marien, I., Vermeulen, B., Lambert, P., Van de Walle, R., Demeester, P.: Assessing the perceptual influence of H.264 SVC Signal-to-Noise Ratio and temporal scalability on full length movies. In: International Workshop on Quality of Multimedia Experience, QoMEx 2009, pp. 29–34 (2009)
11. Waltl, M., Timmerer, C., Hellwagner, H.: A test-bed for quality of multimedia experience evaluation of Sensory Effects. In: International Workshop on Quality of Multimedia Experience, QoMEx 2009, pp. 145–150 (2009)
12. Moebs, S., McManis, J.: A Learner, is a Learner, is a User, is a Customer So what Exactly do you Mean by Quality of Experience? In: 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (2008)
13. El Kaliouby, R., Robinson, P.: Mind Reading Machines: Automated Inference of Cognitive Mental States From Video. In: IEEE International Conference on Systems, Man and Cybernetics (2004)
14. Crane, E.A., Gross, M.: Motion Capture and Emotion: Affect Detection in Whole Body Movement. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 95–101. Springer, Heidelberg (2007)
15. Chanel, G., Kronegg, J., Grandjean, D., Pun, T.: Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals. In: Gunsel, B., Jain, A.K., Tekalp, A.M., Sankur, B. (eds.) MRCS 2006. LNCS, vol. 4105, pp. 530–537. Springer, Heidelberg (2006)
16. Russel, J.A.: A Circumplex Model of Affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980, 2006)
17. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International Affective Picture System (IAPS): Technical Manual and Affective Ratings. NIMH Center for the Study of Emotion and Attention (1997)
18. van Galen Last, N., van Zandbrink, H.: Emotion Detection Using EEG Analysis. Delft University of Technology (2009)
19. up-to-us.rd.francetelecom.com
20. Zawadzki, B., Strelau, J.: Formal Characteristic of Behavior The Temperament Questionnaire (FCZ-K). Faculty of Psychological Tests of the Polish Psychological Society, Warsaw (1997)

# Spanning Trees Structures of Communications Networks

Dounia Lotfi, Mohamed El Marraki, and Driss Aboutajdine

LRIT Associated Unit with CNRST (URAC29), Faculty of Sciences, Mohammed V University, Agdal, B.P.1014 RP, Rabat, Morocco
doun.lotfi@gmail.com, {marraki,aboutaj}@fsr.ac.ma

**Summary.** Graph theory provides a technique for disabling redundant links of a network in order to avoid the switching loop problem. In this paper, we investigate the use of the duality algebraic criteria to find the spanning trees of some switched planar networks. Based on recursive functions, we give new methods that enumerate the spanning trees of some dual graphs. Our theoretical approach guaranties both simplicity and efficiency when compared with the existing approaches and considering the grid network as an application.

## 1 Introduction

A fundamental problem in network communications is disabling redundant links that create loops in switched topologies. Graph theory provides a simple framework to study the communication problems of a given network, such that the vertices represent nodes and the edges represent the links between nodes. A redundant topology is one of the most popular problems that graphs are firm to fix. We assume the topology illustrated in Fig. 1, it is possible that the same packet loop infinitely on a link L, this will create an information redundancy. The aim of this work is to overcome three major problems due to the switching loop, considering the MAC table instability, the broadcast storms and the duplicate frames that will bring a network to an effective obstruction unless prevented. Thus, it seems needful to investigate this redundancy using the graph theory and so then, find the spanning trees of the graphs and the dual graphs that cover the switched networks by maintaining the connection between all of their elements.

Addressing the switching loop problem in redundant topologies requires some basic preliminaries of graph theory. A graph $G$ is a pair of sets $(V, E)$ such that $E \subset [V^2]$, $E$ is the set of edges and $V$ is the set of vertices of $G$. Note that we are dealing with connected planar graphs. In topological sense, the connectivity is the constraint of having a path between each pair of vertices and the planarity provides the intersection between edges only in the vertices

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 283–290.
springerlink.com 

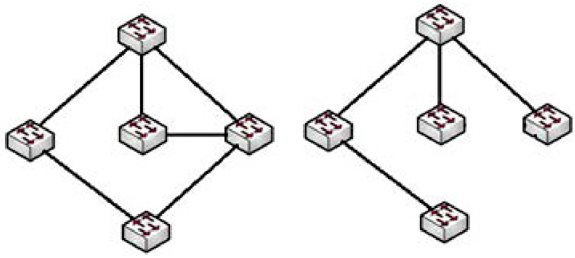


**Fig. 1** Redundant topology and its spanning tree

and forms faces $F$. The issue here is to find the complexity of a network $N$ which is the number of spanning trees of the graph $G$ that illustrates this network and it is denoted by $\tau(G)$. We call a spanning tree of $G$ the subgraph that contains all the vertices and some or all the edges of $G$.

The first method that enumerates the spanning trees of a given graph $G$ was proposed in the Matrix Tree Theorem by Kirchhoff [10] who define the complexity of a graph $G$ as the determinant of its Laplacian matrix (Difference between the graph's degree matrix and its adjacency matrix). Although, the determinant is easy to compute for a given number of vertices using the Matrix Tree Theorem, it will be not convenient to calculate the determinant for huge number of vertices and the Matrix Tree Theorem cannot produce recursive functions that enumerate the spanning trees of a graph $G$. As an alternative, we propose two remedies, considering the illustration of the switched network as a planar connected graph, we prove that the enumeration of spanning trees in a given graph $G$ and its dual is the same. Then, we can use the Kirchhoff Theorem to enumerate the spanning trees of a graph $G$ as for its dual. As a second remedy, we propose recursive functions that count the number of spanning trees of some planar 2-connected dual graphs. Recently, many works succeed in finding the spanning trees of some families of graphs but still challenging task, due to the huge number of nodes in some networks.

As we aforementioned, when graphs are used to study the linkage structure of a network, the nodes are represented by vertices and the links are represented by edges. In topological sense, the dual of a planar graph $G$ is obtained as follows: for every edge $e$ of $G$, we connect the two new vertices in the faces incident with $e$ by an edge $e^*$ crossing $e$. If $e$ is incident with one face, the new vertex in that face will be attached to a loop $e^*$ crossing $e$ again dividing the edge $e$ [8]. In Fig. 2, the dual graph of $G$ is represented by dotted lines and its vertices are drawn by circles, then, the dual of a network $N$ is illustrated by the dual graph $G^*$.

Let $n$ be the number of nodes in the switched network $N$. As a first remedy of the loop link problem, we can search the spanning trees $T$ that cover the

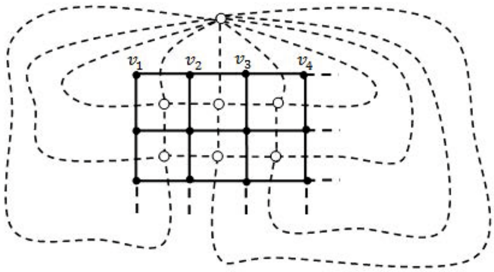



**Fig. 2** Dual graph $G^*$

network $N$ but for a huge number of nodes, it will be more convenient to search the dual spanning trees $T^*$ that contain fewer vertices than $T$.

# 2 Enumeration of Spanning Trees

## 2.1 Spanning Trees of a Dual Graph

There are two methods to enumerate spanning trees that provides the loops removal. The first one is based on the matrix tree theorem given by Kirchhoff, [10] for an arbitrary graph $G$, the number of spanning trees of $G$ is given by the determinant of the Laplacian matrix of $G$. It is known that the major advantage of the Kirchhoff theorem is that the determinant is easy to compute but the downside of this method is that it cannot produce the recursion given the sequence of numbers. We have proposed different methods to derive recursive functions that count the number of spanning trees of some families of graph as remedy of this disadvantage [1], [2] and [4].

The basic idea of this work is to use the matrix tree theorem and the deletion and contraction method that give recursive functions counting the number of spanning trees in order to extend those results to the dual graph $G^*$.

**Theorem 1.** *(The Matrix Tree Theorem) [10] Let $G$ be an indirected graph, the number of spanning trees of $G$ is given by: $L = D - A$ where $D$ is the degree matrix of $G$ and $A$ is the adjacency matrix, then*

$$\tau(G) = det(L_{i,j}(G)) \tag{1}$$

**Theorem 2.** *For a given planar graph $G = (V, E, F)$ and its dual $G^* = (V^*, E^*, F^*)$, the number of spanning trees of $G$ is equal to the number of spanning trees of $G^*$*

$$\tau(G) = \tau(G^*) \tag{2}$$

*Proof.* The construction of a dual graph $G^*$ is achieved as follows: corresponding to each edge $e$ of $G$, we draw an edge $e^*$ which crosses $e$, for every face $f$ of $G$ there is a vertex $v^*$ of $G^*$ and for every vertex of $G$ there is a face $f^*$ of $G^*$. Then, this induces the bijection between the following sets: $F \rightarrow V^*$, $E \rightarrow E^*$, $V \rightarrow F^*$. A spanning tree of $G$ is a tree $A$ that contains the set of edges $E_A \in E$. For each $f$ of $G$, if $deg(f) = n$, then a spanning tree of $G$ contains at most $n - 1$ edges that form this face $f$. One can easily observe that corresponding to each spanning tree of $G$, we can draw a spanning tree of $G^*$ that crosses the edges of $G$ which is not in the spanning tree of $G$. Fig. 3 illustrates a spanning tree of a graph $G$ represented by dotted lines and its corresponding dual spanning tree of $G^*$. The vertices $v^*$ are represented by small white circles and the edges of a spanning tree of $G^*$ are represented by circle dotted lines. Clearly, let $T_G$ and $T_{G^*}$ be the set of spanning trees of $G$ and $G^*$ respectively and $A$ and $A^*$ be the spanning tree of $G$ and $G^*$ respectively, then we have the following bijection: $\sigma : T_G \rightarrow T_{G^*}$, $\mathcal{A} \rightarrow \mathcal{A}^*$, hence $|T_{G^*}| = |T_G|$.

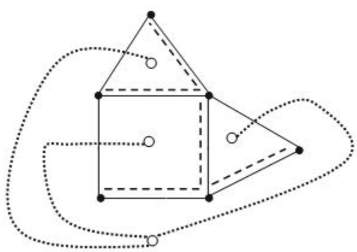


**Fig. 3** Dual spanning tree $A^*$

### 2.2 Spanning Trees of Dual 2-Connected Graphs

Disabling redundant links to avoid switching loops is achieved by searching the spanning trees of a dual graph $G^*$, in this section, we investigate the enumeration of spanning trees for some families of 2-connected planar graphs, we propose recursive functions that provides this counting and we use the deletion and contraction methods studied in our previous works [1], [2] and [4].

A switched network can be mapped by a 2-connected planar graph which has a separation pair $\{x, y\}$, the set $\{x, y\}$ can be two disconnected vertices, an edge or a simple path $p = x, a, b, ..., y$ such that all the vertices have a degree two except $x$ and $y$. The deletion and contraction methods are based on deleting the separation pair $\{x, y\}$ and contracting the two vertices $x$ and $y$, there are denoted by $(G - \{x, y\})$ and $(G.\{x, y\})$ respectively.

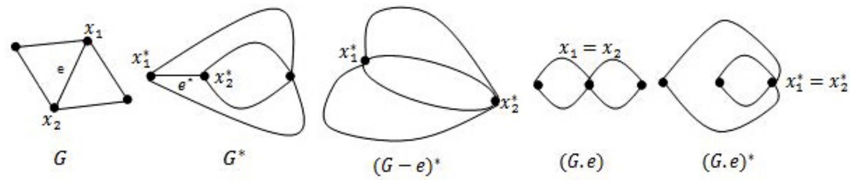



**Fig. 4** Graphs $G$, $(G)^*$, $(G-e)^*$, $(G.e)$ and $(G.e)^*$

**Lemma 1.** *Let $G$ be a planar graph, $e$ an edge of $G$ and $G^*$ is the dual graph of $G$, the operation of deletion and contraction gives us a new graphs which the complexity is given by the following equations:*

$$\tau((G-e)^*) = \tau(G^*.e^*) \quad and \quad \tau((G.e)^*) = \tau(G^* - e^*) \tag{3}$$

*Proof.* Let $G$ and $G^*$ be a planar graph and its dual respectively, $e$ an edge of $G$ and $e^*$ an edge of $G^*$ (see the example in Fig. 4), if we cut an edge $e$ from $G$ we obtain a graph which the dual is denoted by $(G-e)^*$, this later can be also formed by contracting the edge $e^*$ of the dual graph $G^*$ illustrated in Fig. 4, hence $(G-e)^* = G^*.e^*$ then $\tau((G-e)^*) = \tau(G^*.e^*)$. The contraction of the edge $e$ from $G$ (see Fig. 4) give us the dual graph $(G.e)$. One can easily observe that $(G.e)^* = G^* - e^*$ then $\tau((G.e)^*) = \tau(G^* - e^*)$.

**Lemma 2.** *Let $G^*$ be the dual graph of $G$, if $G$ contains simple path $P = x_1, x_2, ..., x_k$ of length $k$, then*

$$\tau((G-P)^*) = \tau(G^*.P^*) \quad and \quad \tau((G.P)^*) = k\tau(G^* - P^*) \tag{4}$$

*such that $P^*$ is the contraction of two ends vertices $x_1^*$ and $x_k^*$.*

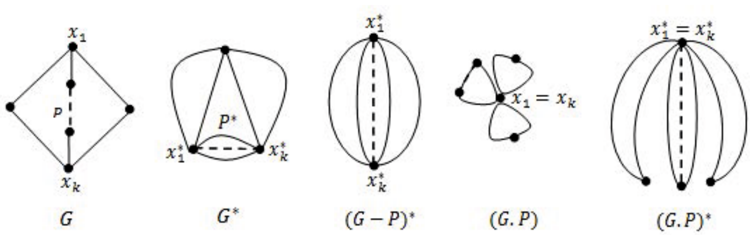



**Fig. 5** Graphs $G$, $G^*$, $(G-P)^*$, $(G.P)$ and $(G.P)^*$

*Proof.* Let $G$ and $G^*$ be a planar graph and its dual respectively, $P$ a simple path of $G$ and $P^*$ is 2-connected subgraph of $G^*$ (see the example in Fig. 5), if we cut the simple path $P$ from $G$ we obtain a graph which the dual is denoted by $(G-P)^*$, this later can be also formed by contracting the two ends vertices $x_1^*$ and $x_k^*$ of the dual graph $G^*$ illustrated in Fig. 5, hence $(G-P)^* = G^*.x_1^*x_k^* \times b_{x_1^*} \times b_{x_2^*} ... \times b_{x_k^*}$ such that $b_{x_i^*}$ are the loops obtained after contracting $x_1^*$ and $x_k^*$. We know that the complexity of a loop is 1 then $\tau((G-P)^*) = \tau(G^*.P^*)$. The contraction of the two ends vertices $x_1$ and $x_k$ from $G$ (see Fig. 5) give us the graph $(G.P)$ which the dual is $(G.P)^*$. One can easily observe that $(G.P)^* = (G^* - P^*) \times P^*$ such that $P^*$ is 2-connected k-cycles, we know that the complexity of k-cycles graph is $k$ then $\tau((G.P)^*) = k\tau(G^* - P^*)$.

**Lemma 3.** *Let $G^*$ be the dual graph of $G$ and $\{x,y\}$ is a separation pair of $G$, we have the following equations:*
*If $\{x,y\}$ is an edge $e$ then $\tau(G^*) = \tau(G^* - e^*) + \tau(G^*.e^*)$.*
*If $\{x,y\}$ is a simple path $P$ then $\tau(G^*) = k(\tau(G^* - P^*) + \tau(G^*.P^*))$.*

*Proof.* Let $G$ and $G^*$ be a planar graph and its dual respectively, $e$ an edge of $G$ and $e^*$ an edge of $G^*$, we know that $\tau(G) = \tau(G-e) + \tau(G.e)$ [4] then from the Lemma 1 we have $\tau((G)^*) = \tau(G^* - e^*) + \tau(G^*.e^*)$. The same goes for the path case.

**Lemma 4.** *Let $G^*$ be the dual graph of $G$, if $G^*$ can be split into two graphs $G_1^*$ and $G_2^*$ through two separation pair $\{u^*, v^*\}$ then the number of spanning trees of $G^*$ is counted as follows:*
*If $\{u,v\}$ is an edge then $\tau(G^*) = \tau(G_1^*) \times \tau(G_2^* - (u^*v^*)) + \tau(G_1^* - (u^*v^*)) \times \tau(G_2^*)$.*
*If $\{u,v\}$ is a simple path then $\tau(G^*) = k\tau(G_1^*) \times \tau(G_2^* - (u^*v^*)) + k\tau(G_1^* - (u^*v^*)) \times \tau(G_2^*)$.*

*Proof.* Let $G$ and $G^*$ be a planar graph and its dual respectively, we have proved that if $e = uv$ is an edge of $G$ then $\tau(G) = \tau(G_1) \times \tau(G_2.(uv)) + \tau(G_1.(uv)) \times \tau(G_2)$ [4] , from the Lemma 1 we have $\tau(G^*) = \tau(G_1^*) \times \tau(G_2^* - (u^*v^*)) + \tau(G_1^* - (u^*v^*)) \times \tau(G_2^*)$. The same proof goes for the path case.

**Lemma 5.** *Let $G^*$ be the dual graph of $G$, $G$ contains simple path $P = x_1, x_2, ..., x_k$ of length $k$, if $G$ can be split into two subgraphs $G_1$ and $G_2$ then $G^*$ can be also decomposed into $G_1^*$ and $G_2^*$. The number of spanning trees of $G^*$ is given by the following equation*

$$\tau((G)^*) = \tau(G_1^*) \times \tau(G_2^*) - k^2\tau(G_1^*.P^*) \times \tau(G_2^*.P^*). \tag{5}$$

*such that $P^*$ is the contraction of two ends vertices $x_1^*$ and $x_k^*$.*

*Proof.* Let $G$ and $G^*$ be a planar graph and its dual respectively, $P$ is a simple path of $G$ of length $k$. if $G$ can be split into two subgraphs $G_1$ and $G_2$ then $G^*$ can be also decomposed into $G_1^*$ and $G_2^*$ the duals of $G_1$ and $G_2$. We know that $\tau((G)) = \tau(G_1) \times \tau(G_2) - k^2\tau(G_1 - P) \times \tau(G_2 - P)$, from the Lemma 2 we have $\tau((G)^*) = \tau(G_1^*) \times \tau(G_2^*) - k^2\tau(G_1^*.P^*) \times \tau(G_2^*.P^*)$.

## 3 Spanning Trees of Dual Grid Networks

As we mentioned above, graph theory provides a technique to overcome the switching loop problem. Many works showed that the simple method to figure out this problem is to model the switched network by a spanning tree that contains all its nodes but not necessarily all the links. So, then we can remove any link with maintaining the connection between all the elements of the network. The enumeration of spanning trees seems difficult for some families of graphs, the goal of this work is to facilitate this count by using the recursive functions given for the dual graphs.

Let $N$ be a network mapped by the graph $G$ such that we define a hierarchical vertex which has multiple connections with all the other neighbours. The other vertices are connected to the nearest neighbour (see Fig. 6). Based on the geometrical structure of the network $N$, we are not able to derive the recursive function that count the number of spanning trees in $G$, we propose as remedy to search the spanning trees of $G^*$. Fig. 6 illustrates the graph $G$ that models $M$ and its dual $G^*$ called the grid graph.

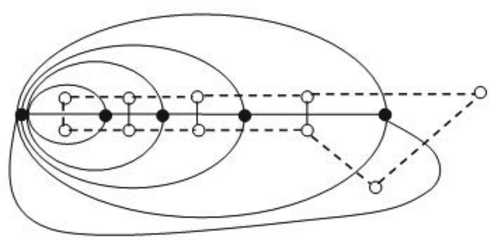


**Fig. 6** The dual of grid network

The complexity of the grid graph is given by: $\tau(G_n^*) = (\frac{1}{2} + \frac{1}{\sqrt{3}})((2 + \sqrt{3})^n - (2 - \sqrt{3})^{n+2}), \quad n \geq 1$ [4], [2]. Then, from the Lemma 1 the number of spanning trees of the network $N$ is given by $\tau(G_n) = (\frac{1}{2} + \frac{1}{\sqrt{3}})((2 + \sqrt{3})^n - (2 - \sqrt{3})^{n+2}), \quad n \geq 1$.

## 4 Conclusion

Graph theory provides a modelling that disables the loop links in the switched networks. In this paper, we used the algebraic planarity criteria to prove that the enumeration of spanning trees in a graph $G$ and its dual is the same using the Matrix Tree Theorem. Based on the deletion and contraction method, we derive recursive functions counting the number of spanning trees of some dual graphs. Using the method cited previously, We have proved that the enumeration of spanning trees of a graph $G$ returns to calculate the number of spanning trees of the dual graph $G^*$ . Our method yields good theoretical results that provide the counting of spanning trees in a 2-connected planar graph $G$ and its dual which we are not able to compute using the classical method. Our future work takes two directions, we will propose a spanning tree algorithm based on the deletion and contraction method in order to derive recursive functions searching the spanning trees of 2-connected planar graphs and we will investigate the family of k-connected planar graphs, in particular, the classical one, in order to extend our results.

## References


1. Lotfi, D., El Marraki, M., Modabish, A.: Recursive relation for counting the complexity of butterfly map. Journal of Theoretical and Applied Information Technology 21(1), 43–46 (2011)
2. Lotfi, D., El Marraki, M., Aboutajdine, D.: Spanning tree recursions for crosses maps. Journal of Theoretical and Applied Information Technology 31(1), 1–7 (2011)
3. Myers, B.R.: Number of spanning trees in a wheel. IEEE Transactions on Circuit Theory CT-18, 280–282 (1971)
4. Modabish, A., Lotfi, D., El Marraki, M.: The number of spanning trees of planar maps: theory and applications. In: Proceeding of the International Conference on Multimedia Computing and Systems IEEE, ICMCS 2011, Ouarzazate, Morocco, pp. 1–6 (2011)
5. Modabish, A., El Marraki, M.: The number of spanning trees of certain families of planar maps. Applied Mathematical Sciences 5(18), 883–898 (2011)
6. Haghighi, M.H.S., Bibak, K.: Recursive relations for the number of spanning trees. Applied Mathematical Sciences 3(46), 2263–2269 (2009)
7. Bogdanowicz, Z.R.: Formulas for the number of spanning trees in a fan. Applied Mathematical Sciences 2(16), 781–786 (2008)
8. Nishizeki, T., Saidur, R.M.: Planar Graph Drawing. Lecture Note Series on Computing, vol. 12. World Scientific Publishing Co. Pte. Ltd, Singapore (2004)
9. Desjarlais, M., Molina, R.: Counting spanning trees in grid graphs. Congressus Numerantium 145, 177–185 (2000)
10. Kirchhoff, G.: Über die auflösung der gleichungen auf. Welche man bei der Untersuchung der Linearen Verteilung Galvanischer Ströme geführt Wird. Ann. Phy. Chem. 72, 497–508 (1847)

# Online Social Networks: Emerging Security and Safety Applications

Michał Choraś[1,2], Marco Manso[3], Damian Puchalski[1],
Rafał Kozik[2], and Krzysztof Samp[1]

[1] ITTI Sp. z o.o., Poznań, Poland
michal.choras@itti.com.pl,damian.puchalski@.itti.com.pl
[2] University of Technology and Life Sciences, Bydgoszcz, Poland
chorasm@utp.edu.pl
[3] TEKEVER Communication Systems, Lisbon, Portugal
marco.manso@tekever.com

**Summary.** In this paper we discuss emerging security and safety applications of Online Social Networks (OSNs). Due to high worldwide popularity social networking can be considered not only as an effective communication channel and efficient information sharing platform, but also as a useful mechanism to enhance public safety and security, including crisis response and recovery efforts, urban operations, homeland security, and mass events planning. Therefore, in this paper we present real examples of how OSNs can be used for those purposes. We conclude by proposing a correlation framework that can effectively increase capabilities of OSN data management and processing for public safety.

## 1 Introduction and Motivation

The rapid evolution of information systems and mobile technology enabled a wide variety of users to easily create and share information through emerging social networks (e.g., Facebook, Twitter, MySpace and LinkedIn) anywhere and anytime. Additionally, websites hosting multimedia content have extensively evolved towards providing online services where rich information (e.g. high-quality imagery and video)can be easily uploaded by and exchanged with the wide public (e.g., MySpace, YouTube and Livelink). There are also solutions targeting users who want to maintain and share online diaries (Blogger, Twitter, Wordpress). All these technologies allow users to share knowledge, share emotions and stay in touch with their friends and families. At the same time, institutions and companies also regard this as an opportunity to exploit such information for their specific needs and purposes.

In this paper we analyze the possible usage of OSNs (Online Social Networks) by public safety agencies, such as law enforcement, the Police and crisis response. We also describe current aspects related to OSN data analysis

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 291–302.
springerlink.com 

and suggest a correlation approach to improve decision making including OSN data.

The paper is structured as follows: in Section 1.1 current market share and status of OSN in Europe is presented. Section 1.2 is devoted to detailed description of OSNs in Poland. Section 2 describes emerging applications of OSNs, such as urban management, crisis management, law enforcement and mass events planning. In Section 3 a correlation approach for OSN data analysis is suggested. Conclusions are given thereafter.

### 1.1 Worldwide Information and Social Media Growth

Worldwide social networking penetration in 2011 was estimated at above 80% of the Internet users in the 15-54 age segment. In Europe, Facebook is currently the leading social networking platform. The highest values of FacebookŠs market share can be observed in Turkey (90.4%) and the UK (81.7%). It is estimated that in USA, Facebook is visited by 58% of all online adults at least one a month [1]. The second most popular social service in Europe is Twitter. Twitter offers a particular set of features that have led to its success, including its short message approach (a maximum of 140 characters per tweet) ensuring an efficient tweets-per-second (TPS) capacity: the record was recently set at 7196 TPS, during the final game of the WomenŚs Football World Cup Final 2011 (17th July 2011) [2]. Relating to crisis response efforts, the Japanese earthquake in March 2011 prompted a total of 5530 TPS [3] (ranked 4th, in terms of the highest number of TPS).

### 1.2 Online Social Network Popularity in Poland

According to the Net Track research conducted by MillwardBrown SMG/KRC in August 2010, the Internet was used by 52.8% of Poles - about 20.1 million people. The number of Internet users that participate in discussions on message boards or chats, write blogs, or use social networking portals as social communities users was estimated at 11.6 million - 58% of the total number of Polish Internet users, based on the results of Net Track. In 2010, over 92% of Polish Internet users at least once visited the NK.pl portal, and 83% created their own profile using this service. Almost 75% of respondents visited the country's second most popular OSN Ű Facebook.com, and three out of five Polish Internet users were registered and had an account there. In 2011, Facebook, and NK had over 11 million users in Poland [4]. Fig.1 shows the number of social networks users in a period of one year and highlights the fact that Facebook became the most popular social platform in Poland in September 2011. According to Social Media report [5], more than 75% of respondents declare that the main motivation for using social media is maintaining contacts with friends. More than 60% Internet users use these sites to maintain relationships with friends from work or school, and to view and comment content published by other users. 33% of the respondents visit social networking sites to meet new friends or to acquire updated information from various sources. In 2011, users of NK.pl sent almost 8 private messages (most of them via NKtalk, a dedicated instant messaging software), uploaded 360 thousands photos and posted over 1.1 million comments about images published by their friends.

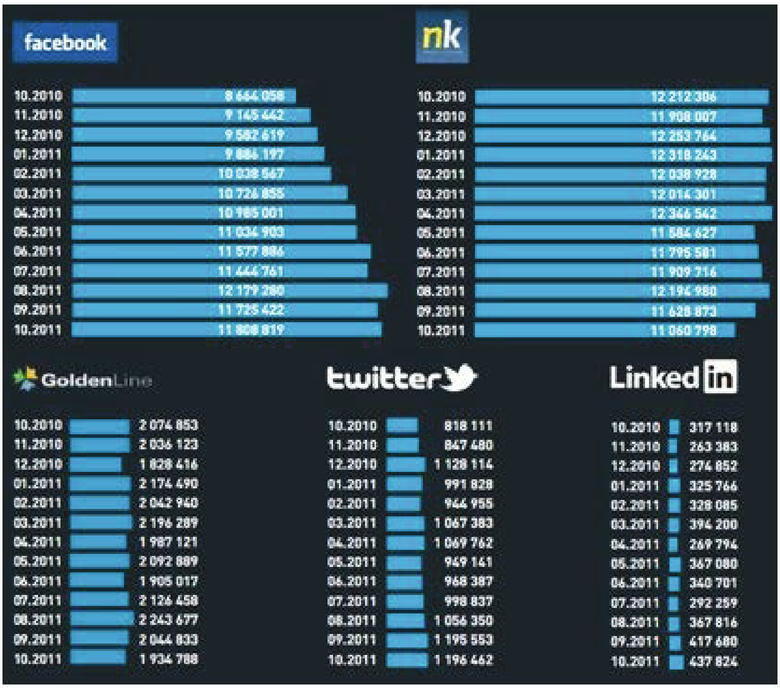



**Fig. 1** Number of real users of the most popular social networks in Poland [4]

A relatively small percentage of Internet users are interested in various forms of entertainment such as games and quizzes and publishing their own posts about the events of life (30, 28 and 27%, respectively) [5]. On the other hand, an upward trend in using of OSNs for marketing, e-commerce (social commerce), geo-localization services and users profiling (for better content targeting) is continuously observed [4].

## 2 Emerging Security and Safety Applications of OSNs

### 2.1 Social Engagement

In 2008UK Police began experimenting with social media, using OSNs as a basis for dialogue with citizens and exploiting growing social media popularity as a means for social engagement. Such activities are supported by the "Engage" strategy [6] regulating all aspects of the police and citizens interactions via OSNs. The "Engage" strategy promotes mainly Twitter platform as the communication channel, due to its worldwide range and popularity, openness and accessibility both for the police officers and the general public, and due to the fact that costs of entry are minimal. However, the strategy envisages also other areas in which capabilities provided by social networking platforms can be beneficial, i.e. as a (open) source of intelligence, real-time information about safety issues, accidents, incidents and for emergency purposes. On the other hand, social media can be considered as a tool for knowledge sharing within policing organizations. They can also provide information helpful in fighting Internet-based crime and protecting the public from harm on the Internet [7].

### 2.2 Urban and Crisis Management

Crisis management can be seriously considered as a potential area of social net-working usage. Social media due to their popularity and ability to broadcast information easily (alerts, warnings, addressing rumors or correcting misinformation) seem to be an adequate option for crisis communication [8][9]. The main features that determine the usefulness of social media (in the crisis management area)include popularity (i.e., number of active users), quickness of information sharing and a growing number of users who tend to use simultaneously many different social networking platforms and other online available tools. The ability of tracking important events (in 24/7 fashion regardless of time zones) as well as increased usage of modern mobile devices for social networking are also considered as beneficial for Future Internet innovative online applications. In the following paragraphs we will describe some real examples of how OSNs have been used in crisis situations and how quickly the information about incidents is disseminated. Specifically, the two cases described below highlight the role of emerging Internet-based services and applications that, due to their innovativeness, can enhance capabilities for crisis management entities, end-users and communities.

<u>Case 1: Emergency Landing at Warsaw Chopin Airport</u>

Alert24 portal [10] is maintained by the publisher of gazetawyborcza.pl and gazeta.pl information portals, as well as of a popular daily newspaper "Gazeta Wyborcza". Alert24 allows users to send messages via MMS, SMS or text, images and videos through an online form. These messages are often published before any official confirmation of a given incident by reporters/news agencies is available, thus being the first information source about a disaster or an accident. A similar approach is used in the Kontakt24 portal, a subsection of tvn24.pl [11] news website (on-line edition of nationwide information television channel - TVN24).

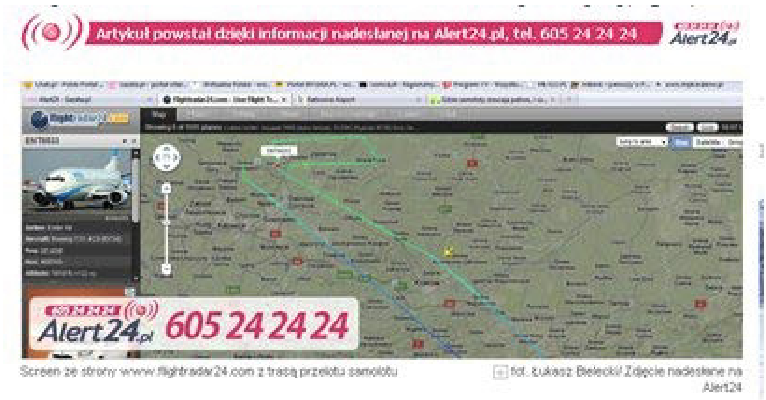



**Fig. 2** Screen from the www.flightradar24.com posted by an Internet user on 8th November 2011 during the air incident in Katowice


**Fig. 3** Headline of the first official news report about incident (1.5h after unofficial alerting by Internet-user)

Recent air incidents in Poland show the benefits (such as rapid information distribution) of publishing news through such portals. On November 1st2011, at 14:40 CET, Boeing 767 (flight from New York to Warsaw) successfully performed an emergency landing at Warsaw Chopin Airport. Polish public TV information channel showed the moment of landing during a live broadcast (only a few second before landing [12]), while Internet users sent images of Boeing (escorted by military fighter jets), circling over the airport and consuming excess fuel [13]. These images had been published at least several minutes before the TV live broadcast and were the first multimedia resources related to the incident and available to a wider audience.

Case 2: Emergency landing at Katowice Airport

On 8 November, chartered aircraft Boeing 737 (flying from Katowice to Egypt) turned back from the area of Polish-Slovak border and circled about

1 hour over the airport in Katowice. An Internet user tracked the flight path through the flightra-dar24.com portal, and sent this information to the Alert24 portal [14] (Fig.2).

The aircraft landed safely about 12:00 after excess fuel had been consumed. The first information about this incident was delivered via the Alert24 portal when the plane was still in the air (before 12:00). After receiving the message, reporters of gazeta.pl confirmed with the press officer of the Katowice airport that the aircraft had technical problems. However, official and confirmed news about the described incident appeared on information portals(including gazeta.pl) later on, after 13:00 [15] (Fig.3).

The two examples mentioned show that citizens who have access to video cameras, mobile phones and access to Internet (including social networks) are often the first sources of information. Quickness of such information publication (after immediate confirmation by reporters) can be critical and invaluable during the crisis incident occurrence. Such conclusion can also be drawn from the Oslo shooting on $22^{nd}$ July 2011. In this tragic situation, typical communication mechanisms, e.g. 112, lacked capacity to handle requests. On the other hand, online social networks handled hundreds of requests per minute (or seconds, if considering Twitter). This proves the capabilities of emerging OSN-related applications for efficient information sharing, management and analysis.

### 2.3 Mass Events Planning

The following two cases can be considered as a motivation for the idea of correlation of information from OSNs. The following examples of demonstrations prove that analysis of online social networking sources, correlation of information and adequate reasoning could be a key factors in crime prevention and protection of mass events.

**Warsaw case (11th November 2011)**

For several weeks, a coalition of leftists, anarchists, proabortionists, Greens and gayrights activists had been publicizing plans to block the Independence March being organized by nationalist youth groups All-Polish Youth and the National Radical Camp on 11th November 2011 in Warsaw. Therefore, the Independence March was secured by c.a. 3000 policemen, however forces designated to protect the event were unable to prevent riots. As a result, more than a dozen people, including nine policemen, were injured in the clashes, and rightist demonstrators later set fire to a television van covering the incident [16]. During the demonstration, 210 people were arrested as a result of clashes with the police. The city of Warsaw and the Police estimated losses were 20k€ and above 60k€[17], respectively.

**Poznań case (19th November 2011)**

Eight days after the incidents in Warsaw, Poznań Police were advised by Warsaw Police not to ignore the threat related to Poznań Equality March. The Police were aware that various online forums and other public sources of information (e.g. social networking) indicated plans to disturb the march, to provoke and confront of the opposite sides [18]. As a result, the Police forces were deployed along the entire route of the march (5km), separating both sides and preventing a possible escalation of aggression and crime incidents [19]. Therefore, analysis of open intelligence sources allowed the police authorities to develop an adequate plan and to involve an appropriate number of staff to protect the event.

### 2.4 OSNs for Law Enforcement

Rapid growth of Online Social Networks (OSNs) has created additional capabilities for law enforcement and security services (e.g. the Police, law enforcement and special agencies, services responsible for airport security, etc.). For example, in addition to traditional purposes of social media usage focused on usersŠ information exchange, social networks are being increasingly exploited by the Police for their own operational purposes. Platforms like Facebook and Twitter can help law enforcement officials to quickly share information with a large audience. Instead of waiting for the local news channel to cover a story, police departments can instantly post important messages online generating real-time alerts to subscribed users. In addition to problem of sharing urgent updates, social media engage the community, and encourage better interaction between law enforcement and civilians. It is worth noticing that some Police Departments also use their own platforms to reach citizens (e.g. North Yorkshire PD in the UK). However, the most innovative effect of social media exploitation by law enforcement comes from the new ways that the Police use platforms like Facebook, YouTube and Myspace to support investigations [20]. One of the most recent examples of OSN data analysis for surveillance and law enforcement purposes is the case of two British tourists that were detained on 23rdof January 2012 by the personnel of The Department of Homeland Security, after joking on Twitter that they were going to 'destroy America' and 'dig up Marilyn Monroe' (Fig.4). Leigh Van Bryan was handcuffed and kept under armed guard in a cell for 12 hours after landing in Los Angeles with his friend Emily Bunting. The Department of Homeland Security flagged him as a potential threat when he posted the message on Twitter to his friends about his planned trip to Hollywood. The passports of British tourists were confiscated and they were interrogated despite the fact that they tried to explain that term 'destroy' was British slang for 'party'. As a result, they were held on suspicion of planning to 'commit crimes'. Van Bryan was also quizzed about another tweet about "diggin' Marilyn Monroe up!" which quoted a popular US comedy Family Guy.


**Fig. 4** Tweet classified as a threat

### 2.5 OSN Data from the Military Perspective

Online Social Networks are also monitored by the army and militaryrelated agencies in order to protect their forces (such as ISAF in Afghanistan) and to prevent acts of terror. The main threat coming from the use of OSN applications is the fact that information posted by a user on the Internet can be accessed by anyone, including her/his potential adversaries. Terrorists may also use OSNs (and the web) in an attempt to reach a wide audience of people, spread propaganda through social web, find potential recruits or even conduct a research on governmental and military assets using the information available online. Fig.5 shows four generic potential activities of an adversary

These activities are listed below and a short explanation is provided:

- Planning: the World Wide Web enables the exchange of information via email, instant messaging or posting topics on social web sites. Services like Facebook or Twitter may be used by adversaries/terrorists to plan their activities and exchange information. Their interest in these communication mechanisms stems from their efficiency: e.g., current mobile phones implement applications which notify the user when a new thread has been added to a Twitter or Facebook account. Thus a wide range of users from various geographical locations can be reached instantly.
- Research: the unawareness of threats of social web users creates another point of vulnerability to governmental and military organizations, especially in case where incautious government or military employees are posting too much information on the public web (e.g. about themselves or the

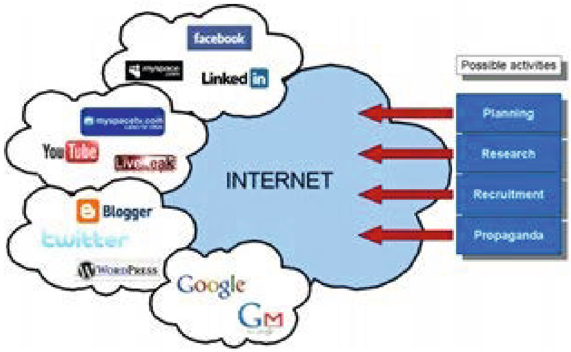



**Fig. 5** Four potential activities that may be a threat for homeland security

missions they are involved). Such information can be used against a specific person as well as the organization employing them. In fact Google and social web services can be comprehensive tools for adversaries to gather information about governmental and military assets or to track down a particular person.

- Recruitment: the advanced search engines built in social web portals provide a comprehensive tool for adversaries to track down specific persons. In particular this feature can be used to recruit new troops or extend the adversaryŠs range of influence. Incidentally, Twitter was used during the Iran elections to recruit people sharing similar political opinions to force a voluntary DoS attack against governmental sites [21]. This example shows how easy it is to recruit people to carry out a cyberattack.
- Propaganda: the fact that information on World Wide Web is publicly available and accessible for users stimulates the use of Internet as the underlying medium to facilitate propaganda. Services as YouTube, Blogger or Twitter are used by adversaries to promote their beliefs and intimidate people by posting links to movies or pictures filled with violence, usually showing human decapitations and executions. Another way to reach the public is the use of "E-flets", described in [22]. Additionally, the adversary may cause deception by distributing false information.

## 3 Proposition of the OSN Data Correlation Framework

The case of Van Bryan and several other cases [23] show that the majority of current digital forensic evidence analysis considers techniques involving only one source of digital evidence during the investigation. However, the example from the Australian New South Wales Police E-Crime Unit [24] shows that such investigations can be optimized and improved (in terms of reliability and effectiveness) through involvement of such means as hard disk drives, mobile phones, personal digital assistances and numerous other consumer sources of data. This highlights a need for developing and evaluating new techniques for visualization and analysis(also correlation of data techniques)of such diverse and heterogeneous sources of information. Event correlation is a powerful computational paradigm for real time processing of continuous data streams. It focuses on producing a single information item from a large number of events and recognizing rare, typically crucial situation that are impactful for monitored system [26]. The event correlation paradigm is currently widely investigated in different domains such as business processes monitoring [27] and network behavior analysis [28][29][30], while the research concerning the OSNs is in our opinion still strongly focused on modeling, characterizing and analyzing the social relations between their users. The more important aspects that refer to content and meaning of the data being posted is still underestimated. Therefore we in this section we propose an novel approach and proposal of framework that aims at correlating and understanding the meaning of scattered data.

Additionally, data management in the described security and safety applications is a complex and difficult task due to the fact that such data is characterized by both structural and semantic heterogeneity. This increases a need for data integration and ingestion in order to support entities responsible for emergency/crisis management during disaster recovery. Data during disaster management may be provided by various, often independent sources of information. Users might be interested in different kinds of knowledge, therefore data mining mechanisms can be considered as capable of enhancing usefulness and reliability of such information.

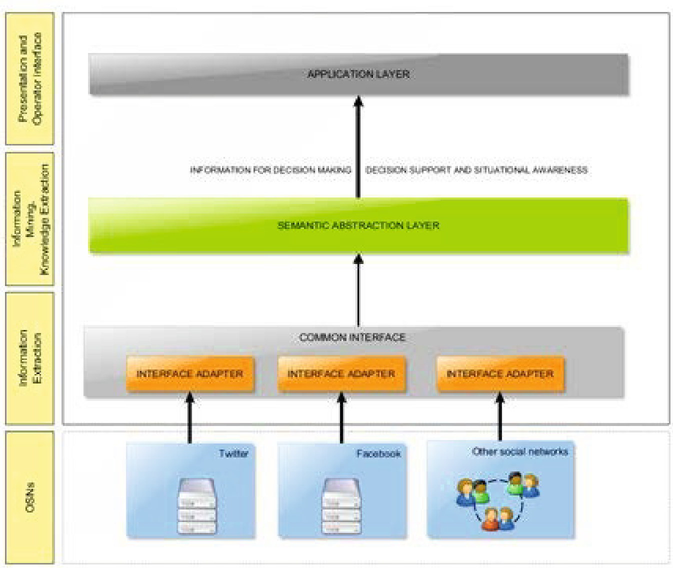



**Fig. 6** OSN data correlation framework

As it is shown in Fig.6, the data obtained from social networks (common interface layer) need to be understood first. In other words, an automated mechanism needs to process the raw text in order to tag it (assign a set of labels that allow document/post classification). Typically, this involves a wide range of mechanisms and algorithms such as pattern recognition for discovering interesting associations and correlations or clustering and trend analysis to classify recognized events for undesired reoccurrences of given phenomena. Once the data sources are correlated and equally annotated (with common semanticbased language) further applications and services such as decision support tools can be deployed Due to the fact that the realworld data in security and safety applications tend to be incomplete, noisy, inconsistent, multidimensional and multisensory etc., development of missing/incomplete data correlation approach in order to increase the situational awareness can be especially beneficial [25].

## 4 Conclusions

The major contribution of this paper is the overview of emerging security and safety applications of OSNs. Moreover, we suggested a correlation approach and framework for OSNs data management and analysis. OSNs are mainly used to post information, maintain contacts with friends etc. However, as we have presented, OSNs can also be successfully used by the Police, law enforcement agencies as well as in the urban management and mass events planning domains. The key challenges are large volume of data (text, images) etc. and quickly changing language and context. Therefore there is a need to develop new tools to automatically handle such tasks. We are currently working on the correlation mechanism that will be able to support humans in large volume data management and in making decisions on the basis of OSNs data in security and safety applications.

## References


1. The 2011 Social Media Consumer Trend and Benchmark Report, Experian Information Solutions (2011)
2. http://www.simplyzesty.com/social-media/twitter/top-10-most-tweeted-about-events-ever-so-far/ (last visited March 12, 2012)
3. http://blog.twitter.com/2011/06/global-pulse.html (last visited March 12, 2012)
4. Internet Standard, Social Media 2012. Raport, II Edycja (2012)
5. Internet Standard, Social Media 2010. Raport, I Edycja (2010)
6. National Policing Improvement Agency, Engage: Digital and Social Media for the Police Service, NPIA, London (2010)
7. Crump, J.: What Are the Police Doing on Twitter? Social Media, the Police and the Public Policy and Internet 3(4) Article 7, (2011)
8. Pring, C.: Using Social Media for Crisis Management (2011)

9. Birgfeld, B.: Why Crisis Management and Social Media Must Co-exist (2010)
10. http://www.alert24.pl/alert24/0,0.html
11. http://kontakt24.tvn.pl/
12. http://tvp.info/modul-specjalny/tvp-info-jako-jedyna-telewizja-pokazala-ladowanie-boeinga-na-zywo-zobacz-je-u-nas/5596559 (last visited March 12, 2012)
13. http://kontakt24.tvn.pl/temat,wyladowal-szczesliwie-eskortowaly-go-f-16,151146.html (last visited March 12, 2012)
14. http://www.alert24.pl/alert24/1,84880,10608235,Awaria-czarterowego-boeinga-Zawrocil-na-lotnisko.html (last visited March 12, 2012)
15. http://wiadomosci.gazeta.pl/wiadomosci/1,114873,10608235.html (last visited March 12, 2012)
16. http://uk.reuters.com/article/2011/11/11/uk-poland-clashes-idUKTRE7AA5BQ20111111 (last visited March 12, 2012)
17. http://www.tvn24.pl/12690,1724324,0,2,policja-wycenila-straty-po-zadymach-11-listopada-w-stolicy,wiadomosc.html (last visited March 12, 2012)
18. http://wiadomosci.gazeta.pl/wiadomosci/1,114873,10668247,Poznan-Nie-dopuscimy-powtorki-11-listopada-w-Warszawie-.html (last visited March 12, 2012)
19. http://www.tvn24.pl/12690,1724857,0,1,marsz-rownosci-i-pikieta-onr-w-poznaniu,wiadomosc.html (last visited March 12, 2012)
20. Sayers, M.: Social Media and the Law: Police Explore New Ways to Fight Crime (2011)
21. http://www.scmagazineus.com/iran-election-protesters-use-twitte-to-recruit-hackers/article/138545/ (last visited March 12, 2012)
22. Carr, J.: Inside Cyber Warfare. O'Reilly Media (2009)
23. Osborne, G., Turnbull, B.: Enhancing Computer Forensics Investigation through Visualisation and Data Exploitation. In: International Conference on Availability, Reliability and Security, ARES 2009 (2009)
24. Beckett, J.: Future Directions Framework. State Electronic Evidence Branch, New South Wales Police, Tech. Rep. (2005)
25. Hristidis, V., et al.: Survey of Data Management and Analysis in Disaster Situations. Journal of Systems and Software 83(10) (2010)
26. Strom, R., Dorai, C., Buttner, G., Li, Y.: SMILE - Distributed Middleware for Event Stream Processing. In: Proceedings of the 6th International Conference on Information Processing in Sensor Networks, pp. 553–554 (2007)
27. Barga, R.S., Caituiro-Monge, H.: Event Correlation and Pattern Detection in CEDR. In: Grust, T., Höpfner, H., Illarramendi, A., Jablonski, S., Fischer, F., Müller, S., Patranjan, P.-L., Sattler, K.-U., Spiliopoulou, M., Wijsen, J. (eds.) EDBT 2006. LNCS, vol. 4254, pp. 919–930. Springer, Heidelberg (2006)
28. Cuppens, F., Miege, A.: Alert Correlation in a Cooperative Intrusion Detection Framework. In: Proceedings of the IEEE Symposium on Security and Privacy, pp. 202–215 (2002)
29. Abadi, D., Carney, D., Cetintemel, U., Cherniack, M., Convey, C., Lee, S., Stonebraker, M., Tatbul, N., Zdonik, S.: Aurora: A New Model and Architecture for Data Stream Management. VLDB Journal 12(2) (2003)
30. Coppolino, L., D'Antonio, S., Elia, I.A., Romano, L.: From Intrusion Detection to Intrusion Detection and Diagnosis: An Ontology-Based Approach. In: Lee, S., Narasimhan, P. (eds.) SEUS 2009. LNCS, vol. 5860, pp. 192–202. Springer, Heidelberg (2009)

# DDoS Attacks Detection by Means of Greedy Algorithms

Tomasz Andrysiak, Łukasz Saganowski, and Michał Choraś

Institute of Telecommunications, University of Technology & Life Sciences in Bydgoszcz ul. Kaliskiego 7, 85-789 Bydgoszcz, Poland
{tomasz.andrysiak,luksag,chorasm}@utp.edu.pl

**Summary.** In this paper we focus on DDoS attacks detection by means of greedy algorithms. In particular we propose to use Matching Pursuit and Orthogonal Matching Pursuit algorithms. The major contribution of the paper is the proposition of 1D KSVD algorithm as well as its tree based structure representation (clusters), that can be successfully applied to DDos attacks and network anomaly detection.

## 1 Introduction

Nowadays cyber attacks are targeted at individual network users but also at critical infrastructures, countries and nations in order to paralyze them. Some successful cyber attacks paralyzed Estonia, Georgia, Iran and most recently Palestine. Such cyber attacks are considered a major threat for critical infrastructures (e.g. power grids, telecommunications systems) and homeland security (e.g. financing system). For example, in 2008 successful DDoS (Distributed Denial of Service) attacks were targeted at Georgian government sites, Georgian president site and servers of National Bank of Georgia. In 2012, hackers from various groups (e.g. Anonymous) launched DDoS attacks that paralyzed Polish public administration and governmental services and websites. Nowadays, cyber attacks are also considered a threat for military networks and critical public administration computer systems.

Emerging new threats and attacks can only be detected by new complex solutions, such as hybrid signature and anomaly based distributed intrusion detection systems. Anomaly detection approach allows for detecting new unknown attacks (so called 0-day attacks) which can not be detected by traditional systems since their signature is yet not known. In this paper we present new $1D\ KSVD$ algorithm for signal-based DDoS detection.

R.S. Choraś (Ed.): Image Processing & Communications Challenges 4, AISC 184, pp. 303–310.
springerlink.com 

## 2 DDoS Attacks: Overview

In recent years Dos and DDos attacks have become a serious problem for various business organizations as well as individual internet users. Their victims are not only the biggest companies from the new technologies sector, which supply mass service, but also powerful governmental organizations from many countries. Regardless huge investment and developed safety procedures, they are usually incapable of effective protection against these kinds of attacks (e.g. as in the case of ACTA related DDoS attacks on Polish public administration websites and services in 2012).

DDos is an upgraded, and most of all, diffused version of the Dos attack. The features concerning an increase in effectiveness and safety of the aggressor have undergone great modification. The typical Dos attack is performed from one computer, while the DDos attack is carried out in a diffuse way (from many previously taken-over computers (bots)). The mentioned computers are most often in different locations. Their may not even be aware that they are participating in an attack on particular resources in the web [1, 2].

There is a number of ways of launching the DDos attacks. First of all, every operating system requires free memory. If the attacker succeeds in allocating the whole accessible memory, theoretically, the system will stop functioning or its performance will drop drastically. Such a violent attack is capable of blocking normal work of even the most efficient IT systems. The second method is based on the use of limitations of the exploited file systems. The third method concerns the use of incorrect operation of network applications, the core of the system or mistakes in configuration of the operating system. Mistakes in TCP/IP stacks can be an instance here. The last method is generating such big network traffic that routers or servers cannot successfully cope with [2, 3]. Typically, the number of generated queries to servers or databases is too large for servers to process and respond.

The problem with DDos attacks lies in the fact that nowadays there are no effective solutions and methods to protect IT systems from them. The only possibility is to limit the results of those attacks by early detection. One of such solutions is detection of network anomalies, which are symptoms of DDos attacks [4].

## 3 Greedy Algorithms Applied to DDoS Attacks and Network Anomaly Detection

Sparse representation is looking for the sparse solution of decomposition coefficients $C$ representing the signal $S$ over the redundant dictionary when the remainder is smaller than a given constant $\varepsilon$, can be stated as:

$$\min \|C\|_0 \ \textit{subject to} \left\| S - \sum_{k=0}^{K-1} c_k d_k \right\| < \varepsilon, \qquad (1)$$

where $\|\cdot\|_0$ is the $l^0$ norm counting the nonzero entries of a vector, $c_k \in C$ represents a set of projection coefficients and $d_k$ are the elements of redundant dictionary D. Finding the optimal solution is an NP-hard problem [5, 6, 7, 8, 9].

A suboptimal expansion can be found by greedy algorithms in means of an iterative procedure, such as the Matching Pursuit algorithm or Orthogonal Matching Pursuit algorithm.

### 3.1 Matching Pursuit Algorithm

The Matching Pursuit (MP) algorithm was proposed in [6]. The aim of the algorithm is to obtain an approximation to the input signal $S$, by sequential selection of vectors from the dictionary $D$. The algorithm follows a greedy strategy in which the basis vector best aligned with the residual vector is chosen at each iteration. Signal $S$ can be written as the weighted sum of these elements:

$$S = \sum_{i=0}^{n-1} c_i d_i + r^n s, \tag{2}$$

where $r^n s$ is residual in an $n$ Ũ term sum. In the first step of Matching Pursuit algorithm, the atom $d_i$ which best matches the signal $S$ is chosen. The first residual is equal to the entire signal $r^0 s = S$. In each of the consecutive $p^{th}$ steps in MP algorithm, the atom $d_p$ is matched to the signal $r^p s$, which is the residual left after subtracting results of previous iterations:

$$r^p s = r^{p-1} s - c_p d_{\varphi_p}, \tag{3}$$

where

$$\varphi_p = arg \max_{i \in \Phi_p} |\langle r^p s, d_i \rangle| \, , \, \varphi_p \in \Phi_p \tag{4}$$

and

$$c_p = \left\langle r^{p-1} s, d_{\varphi_p} \right\rangle . \tag{5}$$

The indices of the $p$ vectors selected are stored in the index vector $\Phi_p = \{\varphi_1, \varphi_2, \ldots, \varphi_{p-1}, \varphi_p\}$, $\Phi_0 = \emptyset$ and the vectors are stored as the columns of the matrix $D_p = \left\{d_{\varphi_1}, d_{\varphi_2}, \ldots, d_{\varphi_p}\right\}$ and $D_0 = \emptyset$. The algorithm terminates when residual of signal is lower than acceptable limit:

$$\|r^p s\| < th, \tag{6}$$

where $th$ is the approximation error.

### 3.2 Orthogonal Matching Pursuit Algorithm

The Orthogonal Matching Pursuit (OMP) algorithm is an improvement of MP algorithm and it was proposed in [10]. Similarly to Matching Pursuit, two algorithms has greedy structure but the difference is that OMP algorithm needs all selected atoms to be orthogonal in every decomposition step.

The algorithm selects $\varphi_p$ in the $p^{th}$ iteration by finding the vector best aligned with the residual obtained by projecting $r^p s$ onto the dictionary components, that is:

$$\varphi_p = arg \max_{i \in \Phi_p} \left| \langle r^p s, d_i \rangle \right|, \varphi_p \notin \Phi_{p-1}. \tag{7}$$

The re-selection problem is avoided with the stored dictionary. If $\varphi_p \notin \Phi_{p-1}$ then the index set is updated as $\Phi_p = \Phi_{p-1} \bigcup \varphi_p$ and $D_p = D_{p-1} \bigcup d_{\varphi_p}$. Otherwise, $\Phi_p = \Phi_{p-1}$ and $D_p = D_{p-1}$. The residual is calculated as:

$$r^p s = r^{p-1} s - D_p \left( D_p^T D_p \right)^{-1} D_p^T r^{p-1} s, \tag{8}$$

where $D_p^T D_p$ is the Gram matrix.

The algorithm terminates when dependency is satisfied (equation 6).

## 4 Structured Dictionaries for Greedy Algorithms

$KSVD$ algorithm has been previously used for 2D signals. Hereby, we propose the new $1D$ $KSVD$ algorithm that we use in anomaly detection task. In particular, we modified the KSVD algorithm proposed in [11].

The main task of $1D$ $KSVD$ algorithm is to find the best dictionary $D$ to represent the signal $S$ as sparse composition, by solving:

$$\min_{D,C} \left\{ \|S - DC\|_F^2 \right\} \textit{ subject to } \forall_i \|c_i\|_0 \leq T, \tag{9}$$

where $\|\cdot\|_F^2$ is the Frobenius norm and $T$ is a fixed and predetermined number of nonzero entries.

The algorithm is divided into two stages:

1. Sparse Coding Stage: Provided D is fixed. We use orthogonal matching pursuit algorithm (mentioned in section 2.2) to compute M sparse coefficients $c_i$ for each sample of signal $S$, by approximation the solution of

$$\min_{C} \left\{ \|s_i - Dc_i\|_F^2 \right\} \textit{ subject to } \|c_i\|_0 \leq T, \ i = 1, 2, \ldots, M, \tag{10}$$

where $s_i$ is a sample of signal $S$.
2. Dictionary Update Stages: Provided both $C$ and $D$ are fixed. We focus on an atom $d_k$ of the dictionary and its corresponding sparse vector $c_T^k$ (i.e. row $k$ of $c_T^k$). The corresponding objective function in equation 9 can be written as:

$$\|S - DC\|_F^2 = \left\| S - \sum_{j=1}^{K} d_j c_T^j \right\|_F^2 =$$

$$= \left\| \left( S - \sum_{j \neq k} d_j c_T^j \right) - d_k c_T^k \right\|_F^2 = \left\| E_k - d_k c_T^k \right\|_F^2 , \qquad (11)$$

where $E_k$ indicates the representation error of the sample of signal after removing the $k^{th}$ atom and its fixed.

The method can be enhanced and improved by generation Of The Dictionary Tree Structure (Clustering).

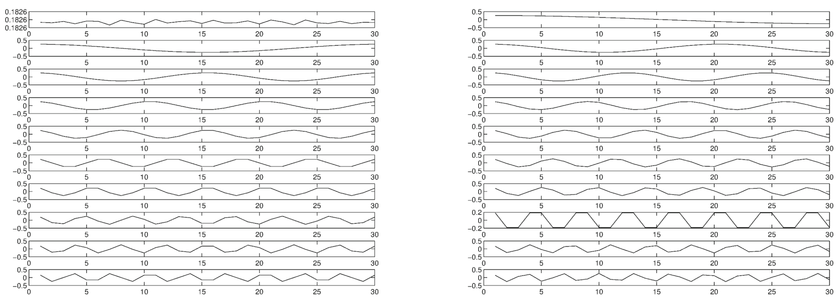


**Fig. 1** Real atoms selected by 1D-KSVD algorithm

Such method creates clusters in the initial dictionary and organizes them in a hierarchical tree structure[12, 13]. The elements from the initial dictionary form the leaves of the tree. As the result, the clustering process produces levels of tree structure. Each branch of the tree has $N$ children and it is fully characterized by the list of the atom indexes $W_n$. A centroid $g_n$ is assigned to the branch of tree that represents atoms of the dictionary in the corresponding subtree.

## 5 Experimental Results

Performance of our approach to signal-based anomaly detection using greedy algorithms and the performance of the presented algorithms was evaluated with the use of CAIDA(2006-2009) and MAWI trace bases [14, 15, 16].

The test data contains attacks that fall into every layer of the $TCP/IP$ protocol stack [17].

For the experiments we chose 10 and 20 minutes analysis windows because most of attacks (about 85%) ends within this time periods [18]. We extracted 15 traffic features in order to create 1D signals for 1D K-SVD and Matching Pursuit analysis.

**Table 1** Matching Pursuit Mean Projection parameter for TCP trace with DDoS attacks (20 min. analysis window) (windows with attack are marked by bold)

| TCP trace (packet/second) CAIDA [15] | Window1 MP-MP (KSVD) | Window2 MP-MP (KSVD) | Window3 MP-MP (KSVD) | MP-MP (KSVD) for trace | MP-MP (KSVD) normal trace |
|---|---|---|---|---|---|
| Backscatter 2008.11.15 | 250.55 | **524.75** | **800.63** | 524.85 | 217.95 |
| Backscatter 2009.10.16 | 272.55 | **532.75** | **780.63** | 541.85 | 220.32 |
| Backscatter 2008.08.20 | **399.68** | 160.79 | 180.54 | 246.60 | 184.50 |
| Backscatter 2009.09.23 | **380.23** | 163.45 | 175.27 | 254.54 | 192.24 |

**Table 2** Matching Pursuit Energy parameter for TCP trace with DDoS attacks (20 min. analysis window)

| TCP trace (packet/second) CAIDA [15] | Window1 $\Psi_{(k)}$ (KSVD) | Window2 $\Psi_{(k)}$ (KSVD) | Window3 $\Psi_{(k)}$ (KSVD) | $\Psi_{(k)}$ (KSVD) for trace | $\Psi_{(k)}$ (KSVD) normal trace |
|---|---|---|---|---|---|
| Backscatter 2008.11.15 | 1.61e+5 | **8.38+05** | **1.01e+6** | 4.13e+5 | 1.72+05 |
| Backscatter 2009.10.16 | 1.53e+5 | **7.23+05** | **1.12e+6** | 4.26e+5 | 1.62+05 |
| Backscatter 2008.08.20 | **2.79e+5** | 2.93e+4 | 6.12e+4 | 1.96e+5 | 4.51e+4 |
| Backscatter 2009.09.23 | **2.91e+5** | 2.73e+4 | 6.03e+4 | 1.85e+5 | 4.41e+4 |

**Table 3** Matching Pursuit Energy parameter for TCP trace (packet/second) with DDoS attacks (20 min. analysis window) for Tree $1D$ KSVD dictionary

| TCP trace CAIDA [15] | Window1 $\Psi_{(k)}$ Tree KSVD | Window2 $\Psi_{(k)}$ Tree KSVD | Window3 $\Psi_{(k)}$ Tree KSVD | $\Psi_{(k)}$ Tree KSVD for trace | $\Psi_{(k)}$ Tree KSVD normal trace |
|---|---|---|---|---|---|
| Backscatter 2008.11.15 | 1.71e+5 | **7.32+05** | **1.30e+6** | 4.15e+5 | 1.73+05 |
| Backscatter 2009.10.16 | 1.59e+5 | **7.28+05** | **1.34e+6** | 4.45e+5 | 1.63+05 |
| Backscatter 2008.08.20 | **2.56e+5** | 2.77e+4 | 6.01e+4 | 1.54e+5 | 4.52e+4 |
| Backscatter 2009.09.23 | **2.79e+5** | 2.72e+4 | 6.21e+4 | 1.86e+5 | 4.41e+4 |

We used two parameters:

- Matching Pursuit Mean Projection (MP-MP)

$$MP - MP = \frac{1}{n} \sum_{i=0}^{n-1} c_i. \tag{12}$$

- Energies of coefficients $\left\|c^{(k)}\right\|^2$, residues $\left\|rs^{(k)}\right\|^2$ and energy of dictionary elements $\left\|d^{(k)}\right\|^2$

$$\Psi_{(k)} = \left\|c^{(k)}\right\|^2 + \left\|rs^{(k)}\right\|^2 + \left\|d^{(k)}\right\|^2. \tag{13}$$

Experimental results are presented in Tables 1 - 3.

In each table, coefficients obtained by means of 1D K-SVD are signed as: MPMP (KSVD) and $\Psi_{(k)}$ (KSVD). Additionally, we added the tree structure to dictionary $D$ generated by 1D KSVD algorithm. Energy parameter $\Psi_{(k)}$ was also calculated for tree structure dictionary.

We compared results achieved by dictionary generated by 1D KSVD and Tree structure 1D KSVD dictionary. Results of detection rate $DR$ and false positives rate $FPR$ were comparable.

Tree structure dictionary has one main advantage which comes from the fact that search time process is significantly lower (up to 50%).

Overall Detection Rate (Overall Detection Rate - ODR is calculated for DR and FPR parameter. ODR takes into consideration set of traffic metrics where at the same time FPR is lowest and DR has highest value. ODR is also calculated for different ADS systems presented in [19, 20]) for tree structure dictionary does not change significantly. For presented DDoS detection method with tree structure 1D KSVD dictionary we obtain $DR = 93,20\%$ $FPR = 12,1\%$ for the tested traces [14, 15, 16].

## 6 Conclusions

This paper presents greedy algorithms applied for DDoS attacks detection task. The major contributions are: the proposition of the new 1D KSVD algorithm and its tree-based structure (clustering). The reported experimental results proved the effectiveness of the proposed method in DDoS detection.

## References


1. Mirkovic, J., Reiher, P.: A Taxonomy of DDoS Attack and DDoS Defense Mechanisms. In: ACM SIGCOMM CCR (2004)
2. Champagne, D., Lee, R.B.: Scope of DDoS Countermeasures: Taxonomy of Proposed Solutions and Design Goals for Real-World Deployment. In: 8th International Symposium on Systems and Information Security, SSI 2006 (2006)
3. Bin, X., Wei, C.: Y. He: A Novel approach to detecting DDoS attacks at an early Stage. Springer Science + Business Media LLC (2006)
4. Teodoro, P.G., Verdejo, J.D., Fernandez, G.M., Vazquez, E.: Anomaly-based network intrusion detection: Techniques, systems and challenges. Computer and Security 28(1-2), 18–28 (2008)
5. Troop, J.A.: Greed is Good: Algorithmic Results for Sparse Approximation. IEEE Transactions on Information Theory, 50(10) (2004)
6. Mallat, S.G., Zhang, Z.: Matching Pursuit with time-frequency dictionaries. IEEE Transactions on Signal Processing 41(12), 3397–3415 (1993)

7. Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations. Journal of Constructive Approximations 13, 57–98 (1987)
8. Gilbert, A., Muthukrishnam, S., Strauss, M.J.: Approximation of functions over redundant dictionaries using coherence. In: 14th ACM-SIAM Symposium on Discrete Algorithms (2003)
9. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM Journal of Computation 24, 227–234 (1995)
10. Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: Asilomar Conference on Signals, Systems and Computers, vol. 1, pp. 40–44 (1993)
11. Aharon, M., Elad, M., Bruckstein, A.: K-SVD. An algorithm for designing overcomplete dictionaries for sparse representations. IEEE Trans. on Signal Processing 54, 4311–4322 (2006)
12. Jost, P., Vandergheynst, P., Frossard, P.: Tree-Based Pursuit: Algorithm and Properties. In: Swiss Federal Institute of Technology Lausanne (EPFL), Signal Processing Institute Technical Report, TR-ITS-2005.013 (2005)
13. Choraś, M., Saganowski, Ł., Renk, R., Hołubowicz, W.: Statistical and signal-based network traffic recognition for anomaly detection. Expert Systems: The Journal of Knowledge Engineering (2011), doi:10.1111/j.1468-0394.2010.00576.x
14. WIDE Project: MAWI Working Group Traffic Archive, http://tracer.csl.sony.co.jp/mawi/
15. The CAIDA Dataset (2006-2009), http://www.caida.org/data
16. Defense Advanced Research Projects Agency DARPA Intrusion Detection Evaluation Data Set, http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html
17. DeLooze, L.: Attack Characterization and Intrusion Detection using an Ensemble of Self-Organizing Maps. In: IEEE Workshop on Information Assurance United States Military Academy, West Point, New York, pp. 108–115 (2006)
18. 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)
19. Wei, L., Ghorbani, A.: Network Anomaly Detection Based on Wavelet Analysis. EURASIP Journal on Advances in Signal Processing 2009, Article ID 837601, 16 pages (2009) doi:10.1155/2009/837601
20. Dainotti, A., Pescape, A., Ventre, G.: Wavelet-based Detection of DoS Attacks. In: IEEE GLOBECOM, San Francisco, CA, USA (November 2006)
21. Coppolino, L., D'Antonio, S., Esposito, M., Romano, L.: Exploiting diversity and correlation to improve the performance of intrusion detection systems. In: Proc. of IFIP/IEEE International Conference on Network and Service (2009)
22. Saganowski, Ł., Choraś, M., Renk, R., Hołubowicz, W.: A Novel Signal-Based Approach to Anomaly Detection in IDS Systems. In: Kolehmainen, M., Toivanen, P., Beliczynski, B. (eds.) ICANNGA 2009. LNCS, vol. 5495, pp. 527–536. Springer, Heidelberg (2009)
23. Gabor, D.: Theory of communication. Journals Electrical Enginners 93, 429–457 (1946)
24. Goodwin, M.: Adaptive Signal Models: Theory, Algorithms, and Audio Algorithms. Kluwer, Boston (1998)

# Index