BIOMETRICS,
COMPUTER SECURITY SYSTEMS
AND
ARTIFICIAL INTELLIGENCE
APPLICATIONS
BIOMETRICS,
COMPUTER SECURITY SYSTEMS
AND
ARTIFICIAL INTELLIGENCE
APPLICATIONS

Edited by

Khalid Saeed
Bialystok Technical University, Poland

Jerzy Pejaś
Szcze cin University of Technology, Poland

Romuald Mosdorf
Bialystok Higher School of Finance and Management, Poland

Springer
The book to which I was asked by the editors to write a foreword is an interesting collection of contributions. It presents the extended versions of the authors' works already introduced at the International Multi-Conference on Advanced Computer Information and Security Systems ACS-CISIM 2005. These contributions had already been reviewed once before the conference for presentation at the conference while some of them passed another selection to be prepared in their current versions for this book.

I am convinced the book will be of help to the researchers in the field of Computer Methods in Biometrics, Security Systems and Artificial Intelligence. They would find the contributions of other researchers of real benefit to them.

I would encourage those who have the book in hands to read it.

Professor Andrzej Salwicki

Faculty of Mathematics, Informatics and Mechanics
Warsaw University, Poland
ACKNOWLEDGEMENTS

The editors are thankful to all contributors whose works have proved to be of great interest to all participants of the International Conference ACS-CISIM, which was held in Elk, Poland in summer 2005.

We also are greatly indebted to the Invited Speakers for their really worth listening and reading keynote talks.

The book could not have appeared without the deep devotion and hard effort of the reviewers, to whom the editors and the contributors really feel grateful. Their reviews for both the Conference Proceedings and this Postconference Book were of great benefit especially to the young researchers whose work still needed others’ professional expertise and comments despite the fact that their scientific research output was positively evaluated. Most of the reviewers did proofreading instead of refereeing. We really are proud of having the book contributions reviewed by them.

Therefore, our indebtedness is due to all the following Professors:

1) Abraham Ajith
2) Bagiński Czesław
3) Bartkowiak Anna
4) Bielecki Włodzimierz
5) Bobrowski Leon
6) Choraś Ryszard S.
7) Dańko Wiktor
8) Huntsinger Ralph
9) Jarmolik Vyacheslav
10) Kuriata Eugeniusz
11) Kompanets Leonid
12) Madani Kurosh
13) Mirkowska Grażyna
14) Ochin Evgeny
15) Petrovsky Alexander
16) Piegat Andrzej
17) Rakowski Waldemar
18) Rogoza Walery
19) Salwicki Andrzej
20) Skarbek Władysław
21) Akira Imada
22) Stepaniuk Jarosław
23) Stokłosa Janusz
24) Tadeusiewicz Ryszard
25) Wierzchoń Sławomir

Editors: Khalid Saeed, Jerzy Pejaś, Romuald Mosdorf
INTRODUCTION

This book presents the most recent achievements in the field of a very fast developing Computer Science. It is a very fascinating science, which still encompasses a number of uncovered areas of study with urgent problems to be solved. Therefore, thousands of scientists are dealing with it elaborating on more and more practical and efficient methods. It is likely that their work will soon result in construction of a very effective, artificial computer-brain.

All scientific works presented in this book have been partitioned in three topical groups:

1. Image Analysis and Biometrics,
2. Computer Security Systems,
3. Artificial Intelligence and Applications.

All papers in the book are noteworthy, but especially we would like to draw the reader’s attention to some particular papers beginning from part 1.

**Image analysis and biometrics** is the branch of Computer Science, which deals with a very difficult task of artificial, visual perception of objects and surroundings and problems connected with it. To the most remarkable papers in this part certainly belongs the invited paper of Anna Bartkowiak et al. where the authors present an interesting mathematical model showing their experience in visualization multivariate data. In his invited paper, Ryszard Choras introduces a survey on Content-Based Image Retrieval showing his and others’ last achievements in this field. Three innovative papers on Face Recognition are also given in the same part. The remaining papers outline their authors’ contribution to Speech Analysis, Signature Recognition using Dynamic Time Warping algorithm and hybrid fused approaches for Speech and Speaker Identification.

**Computer Security and Safety** is at present a very important and intensively investigated branch of Computer Science because of the menacing activity of hackers, of computer viruses etc. To the most
interesting papers in this chapter belongs the invited paper of Janusz Stokłosa et al. It contains an excellent overview of experiments in designing S-boxes based on nonlinear Boolean functions. The authors present also their new algorithm for random generation of perfect nonlinear function. Krzysztof Chmiel’s paper concerns also S-boxes, but in contrast to the previous paper, the author discusses the problem of the differential and the linear approximations of two classes of S-box functions. Two other papers relate to PKI services, which can be used for sending sensitive information and for the public key certificate status validation.

The third part of the book Artificial Intelligence contains 15 absorbing papers, five of which are keynotes and invited papers. The keynotes and invited papers presented at or sent to the ACS-CISIM 2005 conference introduce the latest achievements of their authors W. Dańko, G. Facchinetti et al., A. Imada, K. Madani, G. Mirkowska with A. Salwicki (their keynote paper is not included in this book on their request), R. Tadeusiewicz et al. and S. Wierzchoń et al. The new approaches or notes they show in Computer Artificial Intelligence and its applications are really worth making use of. The remaining papers in this part demonstrate the latest scientific results in the works of their authors in different aspects and areas of Computer Science and its wide applications.

The works contained in the presented book will surely enable you, Dear Reader, to keep pace with significant developments in Computer Science.

We wish you a great satisfaction from reading it.

Professor Leon Bobrowski, Dean
Faculty of Computer Science
Bialystok Technical University
Poland

Professor Andrzej Piegat, Dean
Faculty of Computer Science
Szczecin University of Technology, Poland
# TABLE OF CONTENTS

## PART I - IMAGE ANALYSIS AND BIOMETRICS

*ANDRYSIAK Tomasz, CHORAŚ Michał*

**Image Filtration and Feature Extraction for Face Recognition**  
3

*BARTKOWIAK Anna, EVELPIDOU Niki*

**Visualization of Some Multi-Class Erosion Data Using GDA and Supervised SOM**  
13

*BOBULSKI Janusz*

**Wavelet Transform in Face Recognition**  
23

*CHORAŚ Ryszard S.*

**Content-Based Image Retrieval – A Survey**  
31

*KUBANEK Mariusz*

**Method of Speech Recognition and Speaker Identification Using Audio-Visual of Polish Speech and Hidden Markov Models**  
45

*LIVSHITZ Michael, PETROVSKY Alexander*

**Synthesis of Codebooks with Perceptually Monitored Structure for Multiband CELP-Coders**  
57

*MICIAK Miroslaw*

**The Color Information as a Feature for Postage Stamps Recognition**  
69

*RYDZEK Szymon*

**Iris Shape Evaluation in Face Image with Simple Background**  
79

*SAEED Khalid, ADAMSKI Marcin*

**Experimental Algorithm for Characteristic Points Evaluation in Static Images of Signatures**  
89

## PART II - COMPUTER SECURITY SYSTEMS

*BIELECKI Włodzimierz, BURAK Dariusz*

**Parallelization of Standard Modes of Operation for Symmetric Key Block Ciphers**  
101
PART III - ARTIFICIAL INTELLIGENCE AND APPLICATIONS

ADDABBO Tindara, FACCHINETTI Gisella, MASTROLEO Giovanni 185
Capability and Functionings: A Fuzzy Way to Measure Interaction between Father and Child

DAŇKO Wiktor 197
Remarks on Computer Simulations

IGNATOWSKA Bożena, MOSDORF Romuald 207
Analysis of E-learning System Dynamics

IMADA Akira 217
Can a Negative Selection Detect an Extremely Few Non-self Among Enormous Amount of Self Cells?

KARBOWSKA-CHILIŃSKA Joanna 227
An Average Length of Computations of Normalized Probabilistic Algorithms
ŁEBKOWSKI Andrzej, DZIEDZICKI Krzysztof, TOBIASZ Marcin, ŚMIERZCHALSKI Roman, TOMERA Miroslaw
A Marine Environment Simulator for Testing Ship Control Systems in Dangerous Situations

MADANI Kurosh
Image Enhancement and Image Half-toning Using Fluid Particles Physics Dynamics

ONISZCZUK Walenty
Tandem Models with Blocking in the Computer Subnetworks Performance Analysis

PIECH Henryk, PTAK Aleksandra, MACHURA Marcin
Interchangeable Strategies in Games Without Side Payments on the Base of Uncertain Information on Resources

PIEKARSKI Krzysztof, TADEJKO Pawel, RAKOWSKI Waldemar
Properties of Morphological Operators Applied to Analysis of ECG Signals

REJER Izabela
Input’s Significance Evaluation in a Multi Input-Variable System

SZUSTALEWICZ Adam, VASSILOPOULOS Andreas
Calculating the Fractal Dimension of River Basins, Comparison of Several Methods

TADEUSIEWICZ Ryszard, AUGUSTYNIAK Piotr
Automatic Management of Tele-Interpretation Knowledge in a Wearable Diagnostic Device

WIERZCHOŃ Sławomir, KUZELEWSKA Urszula
Evaluation of Clusters Quality in Artificial Immune Clustering System - SArlIS

ZIENIUK Eugeniusz, SZERSZEŃ Krzysztof, BOŁTUĆ Agnieszka
Convergence Analysis of the Boundary Geometry Identification Obtained by Genetic Algorithms in the PIES
PART I

IMAGE ANALYSIS AND BIOMETRICS
Image Filtration and Feature Extraction for Face Recognition

Tomasz Andrysiak and Michał Choras

Image Processing Group, Institute of Telecommunications, University of Technology and Agriculture, Kaliskiego 7, 85-796, Bydgoszcz, Poland.
[andrys][chorasm]@atr.bydgoszcz.pl

Abstract. In the article we propose Gabor Wavelets and the modified Discrete Symmetry Transform for face recognition. First face detection in the input image is performed. Then the face image is filtered with the bank of Gabor filters. Next in order to localize the face fiducial points we search for the highest symmetry points within the face image. Then in those points we calculate image features corresponding to Gabor filter responses. Our feature vectors consist of so called Gabor Jets applied to the selected fiducial points (points of the highest symmetry) as well as the statistical features calculated in those points neighborhood. Then feature vectors can be efficiently used in the classification step in different applications of face recognition.

1 Introduction

Effective and fast face feature extraction and reliable face feature representation is a key problem in many applications. The most important areas involved in implementing good solutions for that problem are: human-computer interaction, face biometrics, interpretation of face expression, face coding and face tracking. Even though there are many known methods of face detection in images, face feature extraction and representation, still the performance of real-time recognition systems, e.g. for biometrics human identification, is not satisfactory.

In general face feature extraction and representation can be appearance based, 2D geometry based or 3D model based. Since it is difficult to achieve reliable invariance to changing viewing conditions (rotation in depth, pose changes) while basing on 2D geometry [1][15] and 3D models techniques [4][10], currently most of the algorithms are appearance based and use PCA or its derivatives ICA and LDA [2][13][19].

Another popular approach, which is based on Gabor Wavelets, is also appearance based, but local features are computed in the specified points as Gabor filtration coefficients (responses). Such approach relies on filtering the face image by the bank of Gabor filters. Then faces can be efficiently represented by the filter coefficients (so called Gabor Jets) calculated in the extracted fiducial (characteristic) points [12][14][20]. It is mainly because Gabor Wavelets are invariant to some degree to affine deformations and homogeneous illumination changes.
Moreover, Gabor Wavelets are good feature extractors and its responses give enough data for image recognition. The recognition performance of different types of features had been compared in literature and it is shown that Gabor Wavelet coefficients are much more powerful than geometric features [22].

In the article we propose to use Gabor Wavelets for face feature extraction and face feature representation. In section 2 face detection by the ellipse fitting is presented. Face image filtration by the set of Gabor filters is covered in details in the section 3. In section 4 we describe fiducial points extraction and we cover Discrete Symmetry Transform. We also describe the algorithm for determining only the most significant symmetry points within the face.

Face feature calculation in those selected symmetry points is performed next. In each point we compute the responses of Gabor filters for different filter parameters. Moreover, in order to improve the effectiveness of face recognition, we also calculate statistical features. Such algorithm of feature extraction together with the face feature representation is presented in the section 5. Then feature vectors are used in the classification step. Experiments, application to face recognition, discussion and conclusion is given in the next sections.

2 Face Detection

Face detection and extraction of the fiducial (characteristic) points of face are the crucial operations before calculating face features. Face detection is usually performed basing on the Hough Transform but other methods such as deformable templates and color skin models are also used [9][17].

We perform the face detection by the ellipse fitting on the input images. We also use color information and skin models to efficiently localize faces [5][9].

The ellipse fitting algorithm is not restricted to fit the face perpendicularly, therefore if the face is rotated under the angle $\alpha$, the ellipse is also under the angle $\alpha$ towards the vertical axis (as shown in Figure 6). In our method the face detection stage is mainly used for fiducial points selection (Section 4.3).

3 Face Image Filtration

The Gabor Wavelets are used for image analysis because of their biological relevance and computational properties. The Gabor filter kernels model similar shapes as the receptive field of simple cells in the primary visual cortex [16].

Those are multi-scale and multi-orientation kernels and each kernel is a product of a Gaussian envelope and a complex plane wave.

We use Gabor Wavelets to extract the facial features as the set of filter responses with determined scale and orientation values.

The responses image of the Gabor filter can be written as a convolution of the input image $I(\tilde{x})$, with the Gabor kernel $\psi_{\mu,\nu}(\tilde{x})$ such as:
\[ R_{\mu,v}(\tilde{x}) = I(\tilde{x}_o) * \psi_{\mu,v}(\tilde{x} - \tilde{x}_o), \] (1)

where vector coordinates \( \tilde{x} \) of the image \( I(\tilde{x}) \) are equal to \( \tilde{x} = (x, y) \) and \( * \) denotes the convolution operator.

Gabor filters \( \psi_{\mu,v} \) (kernels) can be formulated as:

\[
\psi_{\mu,v}(\tilde{x}) = \frac{\kappa_{\mu,v}^2}{\sigma^2} \exp \left( \frac{\kappa_{\mu,v}^2 \tilde{x}^2}{2\sigma^2} \right) \left[ \exp \left( i\kappa_{\mu,v} \tilde{x} \right) - \exp \left( -\frac{\sigma^2}{2} \right) \right]
\] (2)

The parameters \( \mu \) and \( v \) define the orientation and scale of the Gabor kernels and \( \sigma = 2\pi \).

The wave vector \( \kappa_{\mu,v} \) is defined as follows:

\[
\kappa_{\mu,v} = \begin{pmatrix} k_{x\mu,v} \\ k_{y\mu,v} \end{pmatrix} = \begin{pmatrix} k_v \cos \phi_{\mu} \\ k_v \sin \phi_{\mu} \end{pmatrix}, \quad k_v = 2^{\frac{v+2}{2}} \pi, \quad \phi_{\mu} = \frac{\pi}{8} \mu,
\] (3)

where \( k_v \) is the spacing factor between kernels in the frequency domain.

In most cases Gabor wavelets are used at five different scales and eight orientations [20]. Sometimes other configurations e.g. six orientations are also deployed [22]. Hereby, we use eight orientations \( \mu \in \{0,1,\ldots,7\} \) and three scales \( v \in \{0,1,2\} \) as presented in Figure 1 and 2.

In general, \( \psi_{\mu,v}(\tilde{x}) \) is complex, however, in our approach, only the magnitudes are used since they vary slowly with the position while the phases are very sensitive. Graphical representation of Gabor wavelet kernels are shown in Fig. 1.

![Fig. 1. The kernels of Gabor wavelets at three scales (top-down) and eight orientations (left to right).](image_url)
4 Fiducial Points Extraction

In contrast to many methods which use manually selected fiducial points [22], we search for those points automatically. We use Discrete Symmetry Transform for choosing the points with the highest symmetry within the face. In contrast to a proposed method of DST [6][7], we apply DST onto combined image of Gabor directional filtration images.

Extraction of the High Symmetry Points

The face fiducial points extraction is based on the Discrete Symmetry Transform as presented in [6][7]. Hereby we modify the known method of symmetry points detection by applying Gabor filtered images in the first step of the Discrete Symmetry Transformation.

Our application to detect points of symmetry is developed in two steps. The first step is the Gabor Wavelet filtration for proper values of orientations and resolution. Secondly for each point of the gradient image we compute the symmetry value.

The algorithm to compute modified Discrete Symmetry Transform is following:

1. First we filter the image with the Gabor filters for 2 different orientations \((0, \pi / 2)\). Then we add those images and in result of such filtering we obtain the combined image \(O(x, y)\) such as:

\[
O(x, y) = 1 - \sum_{\mu, \nu} R_{\mu, \nu}(x, y).
\]  

(4)

The image \(O(x, y)\) is presented in the Figure 4 (left).

2. Computation of the \(DST\).
The Discrete Symmetry Transform is computed as the multiplication of the filter response image $O(x, y)$ with the image $M(x, y)$ such as:

$$DST(I(x, y)) = O(x, y) \times M(x, y),$$

where:

$$M(x, y) = \sqrt{\frac{1}{n} \sum_{k=0}^{n-1} (M_k(x, y))^2 - \frac{1}{n^2} \left( \sum_{k=0}^{n-1} M_k(x, y) \right)^2},$$

and:

$$M_k(x, y) = \sum_{(p, q) \in \Pi_r} \left| (p - d) \sin \left( \frac{k\pi}{n} + \alpha \right) - (q - e) \cos \left( \frac{k\pi}{n} + \alpha \right) \right| \times I(x, y)$$

for the following parameters:
- $(p, q)$ are the coefficients of each calculated symmetry point,
- $(d, e)$ are the coefficients of the point belonging to the circle $\Pi_r$ with the distance $r$ from the point $(p, q)$,
- $\Pi_r$ is the circle centred in $(p, q)$,
- $r$ limits the size of the neighborhood of each point $(p, q)$,
- $n$ is the number of axial moments with the slope $k\pi / n$ with $k = 0, 1, ..., n$,
- and where $\alpha$ is the angle between the ellipse and the vertical axis of the image.

The final result of the modified DST computation is presented in Figure 5 (left).

Fig. 3. Results of directional Gabor filtration of for two orientations $(0, \pi/2)$. 
Extracted Symmetry Points Thresholding

The computed \( DST(I(x, y)) \) gives the symmetry points on the \( O(x, y) \) image, but not all the symmetry points become our fiducial points. In order to find the most significant of those points we perform the threshold operation according to the following rule [6]:

\[
\text{ThreshDST}(I(x, y)) = \begin{cases} 
1 & \text{if } DST(I(x, y)) > \text{mean} + 3 \times \text{var} \\
0 & \text{otherwise}
\end{cases},
\]

where \( \text{mean} \) and \( \text{var} \) are the mean value and the standard deviation of \( DST(I(x, y)) \), respectively.

The resulting image \( \text{ThreshDST}(I(x, y)) \) is presented in Figure 5 (right).
Symmetry Points Selection

Moreover, in order to select only points within the faces, we take into account only those points which are localized within the area of the ellipse fitted into original input image as described in Section 2. The procedure of selecting fiducial points from the symmetry points is presented in the Figure 6. Finally, we choose \( N \) points with the highest symmetry value within the detected ellipse and those points become our fiducial points. Usually we select 30 - 40 points of the highest value. Then in those points we calculate face features based on Gabor Wavelets responses (Gabor Jets) as presented in the next section.

![Figure 6](image)

**Fig. 6.** The consecutive steps of the fiducial points selection algorithm.

5 Face Feature Extraction and Representation

We calculate the face features only in the \( N \) extracted fiducial points. We usually choose 30-40 fiducial points. Moreover, we use Gabor filters responses for different orientations and scale as feature vector parameters (8 orientations \( \times \) 3 scales). Therefore for each fiducial point we have a vector of 24 parameters. Such response vectors corresponding to face fiducial points are often called Gabor Jets [20]. The number of parameters in the final feature vector is given by:

\[
F_{GJ} = \mu \times v \times N .
\]  

(9)
In order to enhance the effectiveness of standard Gabor Jets approach, we calculate the second feature vector in the extracted points neighborhood $\Omega (k \times k)$. Still we base on the Gabor-based filtration responses. Therefore the second feature vector consists of the following parameters:

1. mean value for $R^{e}_{\mu,\nu}(\bar{x})$ and $R^{o}_{\mu,\nu}(\bar{x})$:
   
   $$sr^{e}_{\mu,\nu}(\bar{x}) = \frac{1}{n} \sum_{x \in \Omega} R^{e}_{\mu,\nu}(x), \quad sr^{o}_{\mu,\nu}(\bar{x}) = \frac{1}{n} \sum_{x \in \Omega} R^{o}_{\mu,\nu}(x)$$  

2. variance for $R^{e}_{\mu,\nu}(\bar{x})$ and $R^{o}_{\mu,\nu}(\bar{x})$:
   
   $$\nu r^{e}_{\mu,\nu}(\bar{x}) = \frac{1}{n} \sum_{x \in \Omega} (R^{e}_{\mu,\nu}(x) - sr^{e}_{\mu,\nu}(\bar{x}))^2, \quad \nu r^{o}_{\mu,\nu}(\bar{x}) = \frac{1}{n} \sum_{x \in \Omega} (R^{o}_{\mu,\nu}(x) - sr^{o}_{\mu,\nu}(\bar{x}))^2$$

3. module and phase $R^{e}_{\mu,\nu}(\bar{x})$:
   
   $$mr^{e}_{\mu,\nu}(\bar{x}) = \sqrt{(R^{e}_{\mu,\nu}(\bar{x}))^2 + (R^{o}_{\mu,\nu}(\bar{x}))^2}, \quad pr^{e}_{\mu,\nu}(\bar{x}) = \arctan \left( \frac{R^{e}_{\mu,\nu}(\bar{x})}{R^{o}_{\mu,\nu}(\bar{x})} \right)$$

4. moments of order $p$ and $q$:
   
   $$mr^{(p,q)}_{\mu,\nu}(\bar{x}) = \frac{1}{n} \sum_{x \in \Omega} (x^p y^q)$$

where: $R^{e}_{\mu,\nu}(\bar{x})$ and $R^{o}_{\mu,\nu}(\bar{x})$ are the even and odd responses of Gabor Wavelets, respectively, and $\Omega$ is the averaging operator.

The second feature vector is given by:

$$F_S = [sr^{e}_{\mu,\nu}, ..., sr^{o}_{\mu,\nu}, ..., \nu r^{e}_{\mu,\nu}, ..., \nu r^{o}_{\mu,\nu}, ..., mr^{e}_{\mu,\nu}, ..., mr^{o}_{\mu,\nu}, ..., pr^{e}_{\mu,\nu}, ..., pr^{o}_{\mu,\nu}, ..., mr^{(p,q)}_{\mu,\nu}, ...]$$

(14)

Such face feature vectors are used in the face recognition step. The recognition is based on simple comparison of the input image feature vector with all the vectors in the database. In the classification step we base on vector distances in feature space and we use City Block Classifier.

6 Experiments and Results

Reliable face feature extraction and representation depends on properly selected parameters in consecutive steps of our method. Firstly, the parameters of Gabor filters have to be properly tuned so that the effects of filtration can be further used in the calculation of the DST. Moreover, Gabor Wavelets also influence the face feature vector as it consists of the Gabor Wavelet responses.
We experimented with the number of scales and our conclusion is that 3 scales $\nu \in \{0,1,2\}$ are sufficient for reliable face representation. We can also add two more resolutions $\nu \in \{3,4\}$ but the filter responses are small and not distinctive within various images. Secondly, we experimented with the selection of parameters in the algorithm of the symmetry points calculation. Results for different selection of the radius value are presented in the Figure 7.

Another important factor in the symmetry points extraction algorithm is the number of symmetry axis. We use 2 axis: vertical and horizontal since in the case of faces such symmetries are most significant.

In our experiments we considered frontal view face images. For such dataset we obtained correct recognition ratio of 86% while comparing only Gabor Jets feature vector $F_{GJ}$, and 90% while comparing combined both vectors $F_{GJ}$ and $F_S$.

![Fig. 7. Points of the highest symmetry calculation for the radius value of 2 (left), 4 (middle) and 8 (right).](image)

7 Conclusions

In the article we presented an efficient method of face recognition based on image filtering, novel method of automated fiducial points extraction and face feature representation. In our work we base on directional Gabor Wavelets and the modified Discrete Symmetry Transform for the extraction of the face fiducial points. Then in those extracted points we apply parameterized bank of Gabor filters in order to calculate features needed for face representation and recognition. However, in contrast to many known methods, thanks to modified DST we extract the fiducial points automatically. The proposed method gives satisfactory results in comparison to other known implementations of EBGM and Gabor Jets approach to face recognition. We obtain satisfactory correct recognition ratios while using combined both feature vectors $F_{GJ}$ and $F_S$ on frontal view face image databases for example FaDab [3]. Further work includes global features calculation and texture parameters calculation in fiducial points ROI neighborhoods.
References

Visualization of Some Multi-Class Erosion Data  
Using GDA and Supervised SOM  

Anna Bartkowiak¹, Niki Evelpidou²  

¹ Institute of Computer Science, University of Wroclaw,  
Przesmyckiego 20, 51-151 Wroclaw Poland, aba@ii.uni.wroc.pl  
² Remote Sensing Laboratory, University of Athens  
Panepistimiou Zoografou, Athens, Greece, evelpidou@geol.uoa.gr  

Abstract. We present our experience in visualization multivariate data when the data vectors have class assignment. The goal is then to visualize the data in such a way that data vectors belonging to different classes (subgroups) appear differentiated as much as possible. We consider for this purpose the traditional CDA (Canonical Discriminant Functions), the GDA (Generalized Discriminant Analysis, Baudat and Anouar, 2000) and the Supervised SOM (Kohonen, Makivasara, Saramaki 1984). The methods are applied to a set of 3-dimensional erosion data containing N=3420 data vectors subdivided into 5 classes of erosion risk. By performing the mapping of these data to a plane, we hope to gain some experience how the mentioned methods work in practice and what kind of visualization is obtained. The final conclusion is that the traditional CDA is the best both in speed (time) of the calculations and in the ability of generalization.

1 Introduction  

We consider the problem of multivariate data visualization when each data vector has a class (group) assignment. Generally, methods of data visualization perform linear or nonlinear mapping to a manifold of lower dimension. Say, this lower dimension is q. The most common visualization uses q = 2. Generally, it is expected that the projection gives us an idea on the shape of the data cloud. Here, we want more: Using the information about crisp group assignment (‘crisp’ is used here in the opposite meaning of ‘soft’), we seek for such a projection (mapping), which shows distinctly differentiation between various groups of the data.  

When intending a graphical visualization of the data, we should ask in first step about the intrinsic dimensionality of the data. It could happen that all the observed variables are generated by some unobserved variables, so called ‘latent variables’ located in a manifold of lower dimension – and we should know it. Thus, we should ask about the intrinsic dimensionality of the analyzed data. We will use for this purpose the correlation integral C(r) and the correlation dimension D introduced by Grassberger and Procaccia [6], [4]. The next question is: What kind of projection or mapping should we use? The most simple method is the classical one using Fishers’ criterion based on the between and
within class variance and yielding so called ‘canonical discriminant variates’ or canonical discriminant functions [2], [5], [7], [12]. The method belongs to the class of linear methods and is referred to as CDA or Fisherian LDA. The method is extendable to the class of nonlinear methods — by use of appropriate transformation of the data. In particular one may use kernel transformations [8], [9], [5], [11].

Using the kernel approach, Baudat and Anouar [2] proposed a non trivial generalization of the canonical discriminant analysis. They called their algorithm GDA (generalized discriminant analysis). It represents the nonlinear discriminant functions.

As an alternative to the nonlinear GDA we will consider also quite a different algorithm, called SOM supervised (SOM_s) and based on a modification of Kohonens’ self organizing map.

In the following, we will show how the mentioned methods work when analyzing a real data set of a considerable size, i.e. about 3 thousands of data vectors. The data set is subdivided into 5 erosion classes. We take for our illustration only 3 variables known as predictors for the erosion risk. In the case of 3 variables it is possible to visualize the data in a 3D plot. For the considered erosion data, the 3D plot shows plainly that the relations between the variables are highly nonlinear; thus nonlinear projection methods might show a more distinctive differentiation among the erosion classes.

The paper is organized as follows. In Sect. 2 we describe the data and their correlation dimension. Sect. 3 explains the accepted Fishers’ criterion of separation between classes and the principles of building canonical discriminant functions (CDA alias LDA). Sect. 4 shows the nonlinear extension of LDA using the kernel approach proposed by Baudat and Anouar. In Sect. 5 we describe briefly the supervised SOM. Finally, Sect. 6 contains some concluding remarks.

2 The Erosion Data

Our interest in a trustful visualisation of subgroups of data originated from the research of erosion risk observed in the Greek island Kefallinia. The entire island was covered by a grid containing 3422 cells. The area covered by each cell of the grid was characterized by several variables. For our purpose, to illustrate some visualization concepts, we will consider in the following only 3 variables: drainage density, slope and vulnerability of the soil (rocks). The values of the variables were rescaled to belong to the interval [0, 1]. Thus, for our analysis, we got a data set containing N=3422 data vectors, each vector characterized by 3 variables. Using an expert GIS system, each data vector was assigned to one of five erosion classes: 1. very high (vH), 2. high (H), 3. medium (Me), 4. low (L) and 5. very low (vL). A 3D plot of the data is shown in Fig. 1. The data set contains a few outliers, which are strongly atypical observations. Two of them will be removed from further analysis.
Fig. 1. Visualization of the Kefallinia erosion data containing N=3422 data points, subdivided into 5 classes of progressing erosion risk. In some parts of the space the data points are much condensed. Two severe outliers are visible top left – they will be dropped in further analysis.

The different classes of the data set are marked by different symbols and/or colours. Looking at the plot in Fig. 1 one may state that, generally, the distribution of the data is far from normality, also far from the ellipsoidal shape. The hierarchy of the classes exhibits a nonlinear pattern. Some parts of the space show a great concentration of the data points, while some other parts are sparsely populated.

The fractal correlation dimension calculated using the Grassberger-Proccacia index [6], [4] equals $D = 1.6039$. This is the estimate of the intrinsic dimension for the considered erosion data (For comparison, we have performed analogous calculations for two synthetic data sets of the same size, generated from the 3D and 5D normal distributions; as expected, we obtained the values $D_3 = 3.0971$ and $D_5 = 4.9781$ appropriately). Thus – the intrinsic dimension of the considered data set is less than 2 and a planar representation of the data is justified.

The data set contained two big outliers. To not confound the effects of the outliers and the effects of the methods, we have removed the outliers from the analyzed set. Next we subdivided the remaining 3400 data vectors into two parts (halves), each
counting $N = 1710$ data items. The first part (labelled samp1) was destined for learning (establishing the parameters of the models), and the second part (samp2) as test. In the next three sections we will show mapping of the data to a 2D plane using three different methods: canonical discriminant functions (CDA alias LDA), kernel discriminant functions (GDA) and the supervised SOM (SOM_s).

3 Canonical Discriminant Functions

We show now the canonical discriminant functions derived from Fishers’ criterion. The method is called sometimes also LDA [5], [2].

The case of the two-class problem. R. A. Fisher proposed to seek for the linear combination ($a$) of the variables, which separates the two indicated classes as much as possible. The criterion of separateness, proposed by Fisher, is the ratio of between-class to within-class variances. Formally, the criterion is defined as the ratio (see, e.g. Duda [6] or Webb [11])

$$JF2 = [a^T(m_1-m_2)]^2 / [a^TS_w a], \quad (2\text{-class problem})$$

where $a$ is the sought linear form, $m_1$ and $m_2$ denote the sample group means, and $S_w$ is the pooled within-class sample covariance matrix, in its bias-corrected form given by

$$S_w = (n_1S_1+n_2S_2) / (n_1+n_2-2).$$

Maximizing the $JF2$ criterion yields as solution the sought linear combination $a$ for the two-class problem.

In the case of the multi-class problem, – when we have $k$ classes, $k \geq 2$, with sample sizes $n_1, ..., n_k$ totaling $N$, and the overall mean $m$ – the criterion $JF2$ is rewritten as the criterion $J_{pk}$, which accommodates the between class and within class variances:

$$J_{pk} = \sum_j n_j a^T(m_j-m) / (a^TS_w a), \quad j=1, ..., k \quad (k\text{-class problem})$$

where $m_j$ denotes the mean of the $j$-th class and $m$ stands for the overall sample mean. The within class variance $S_w$ is evaluated as ($S_j$ denotes the covariance matrix in the $j$th class, $j=1, ..., k$):

$$S_w = (\sum_j n_j S_j) / (N-k).$$

Maximizing the criterion $J_{pk}$ with respect to $a$ we obtain, with accuracy to the sign, $h$ solutions, i.e. $h$ vectors $a_1, ..., a_h$, $h = \min (k-1, \text{rank of } X)$, with $X$ being the data matrix. From these we obtain $h$ canonical variates: $y_j = Xa_j, \quad j = 1, ..., h$, called also canonical discriminant functions. The separateness of the subgroups, attained when considering the transformation yielded by subsequent canonical discriminant variates, is measured by the criterion $J_{pk}$ evaluated as $J_{pk}(a_j)$ and called also $\lambda_j$. For each
vector $a_j$ we obtain its corresponding value $\lambda_j = \lambda(a_j)$ denoting the ratio of the between to the within class variance of the respective canonical variate derived from the vector $a_j$. Thus a big value of $\lambda_j$ indicates a high discriminative power of the derived canonical variate.

For the analyzed erosion data we got $h=3$ vectors $a_1, a_2, a_3$ and corresponding to them 3 values of $\lambda$ equal to $[22.1995 \ 0.7982 \ 0.0003]$. One may notice that the first canonical variate – compared to the remaining ones - has a very big discriminative power, while the contribution of the third canonical variate is practically none.

The projection of the data, when using the first two canonical variates, is shown in Fig. 2. One may notice that the subgroups are quite well separated. One may notice also that the second canonical variate, which – taken alone – has practically no discriminative power, however, when combined with the first variate, helps much in the display, i.e. in distinguishing the classes of erosion risk. We got very similar values of $\lambda$ and very similar displays both for the learning and the testing data sets (i.e. for samp1 and samp2) – thus the method has good generalization abilities.

Fig.2. Projection of the samp1 data using the first two canonical discriminant functions derived from Fisher’s criterion. The very low and very high erosion points keep opposite position, right and left, in the exhibit. The display for the samp2 data looks identical.
4 Nonlinear Projection Using the Kernel Approach

The CDA, described in previous section, considers only linear functions of the variables and is proper when the groups (classes) are distributed elliptically. For our data this is not the case. Therefore, some nonlinear methods might be better for visualizing the class differentiation. A kind of non-linear discriminant analysis, called GDA (Generalized Discriminant Analysis) was proposed by Baudat and Anouar [2]. Their algorithm maps the input space into an extended high dimensional feature space. In the extended space, one can solve the original nonlinear problem in a classical way, e.g., using the CDA. Speaking in other words, the main idea is to map the input space into a convenient feature space in which variables are nonlinearly related to the input space. The fact of mapping original data in a nonlinear way into an extended feature space was met in the context of support vector machines (SVM) see e.g., [5], [8], [9], [11]. The mapping uses predominantly kernel functions. Direct coordinates -- in the extended space -- are not necessary, because the kernel approach needs only computations of so called 'dot products' formed from the original features.

Generally, the mapping reads

$$\Phi: X \rightarrow F,$$

with $X$ denoting the input space (original data), and $F$ the extended feature space, usually of higher dimensionality as the original data space. The mapping $\Phi$ transforms elements $x \in X$ from the original data space into elements $\Phi(x) \in F$, i.e. elements of the feature space.

Statistical and/or pattern recognition problems use extensively cross products (inner products), e.g. for obtaining the within and between group covariance. To calculate them, a special notation of kernel products was invented. Let $x_i$ and $x_j$ denote two elements (row data vectors) of the input data matrix $X$. The kernel function $k(x_i, x_j)$ returns the inner product $\Phi^T(x_i)\Phi(x_j)$ between the images of these inputs (located in the feature space). It was proved that for kernel functions satisfying some general analytical conditions (possessing so called Mercer properties) the kernel functions $k(x_i, x_j)$ can be expressed as simple functions of the inner product $<x_i, x_j>$ of the original vectors. In such a case, we can compute the inner product between the projections of two points into the feature space without evaluating explicitly their coordinates (N denotes the number of data vectors, i.e. the number of rows in the data matrix $X$):

$$k(x_i, x_j) = \Phi^T(x_i)\Phi(x_j) = k(<x_i, x_j>), \text{ for } i,j = 1, \ldots, N.$$

The GDA algorithm operates on the kernel dot product matrix $K = \{k(<x_i, x_j>)\}$ of size $N \times N$, evaluated from the learning data set. The most commonly used kernels are Gaussians (RBFs) and polynomials.

Let $x, y$ be the two (row) input vectors. Let $d = x - y$. Using Gaussian kernels, the element $z = k(<x_i, x_j>)$ is evaluated as: $z = \exp\left(-\frac{d^*d}{\sigma}\right)$. The constant $\sigma$, called kernel width, is a parameter of the model; its value has to be declared by the user.
Baudat and Anouar use as the index of separateness of the constructed projection a criterion, which they call \textit{inertia}. This criterion is defined as the ratio of the between class to the total variance of the constructed discriminant variates. The inertia criterion takes values from the interval $[0, 1]$. High values of inertia indicate a good separation of the displayed classes.

For our evaluation we have used Matlab software implemented by Baudat and Anouar. For $k = 5$ classes we got 4 discriminative variates. The largest values of inertia were noted, as expected, for the first two GDA variates. What concerns the kernel width $\sigma$, we have tried several values: $\sigma = 0.0005, 0.005, 0.05, 0.5, 1, 4, 6.5, 9, 14$. For each value of $\sigma$, the system has been learning using the \texttt{samp1} data, next the established model was tested using the \texttt{samp2} data. Each run (i.e. calculations for one value of $\sigma$) needed about $12$ minutes of computer time (PC, XPHome, Intel® Pentium® 4, Mobile CPU 1.80GHz, 512 MB RAM). The \texttt{samp1} and \texttt{samp2} data were of size $[1710, 3]$. Thus the computations were quite lengthy. Generally, it was stated that for decreasing values of $\sigma$ the classes appeared more and more separated (for values $\sigma = 0.5$ to $14$, the displays were quite similar). As an exemplary exhibit we show here Fig. 3, obtained for $\sigma = 1$. The resulting inertias for variates no. 1-4 are: $[0.968650 \ 0.705236 \ 0.550547 \ 0.257248]$.

![GDA sample1 N=1710 σ=1](image)

**Fig. 3.** GDA using Gaussian kernels with $\sigma = 1$ applied to the \texttt{samp1} data. Horizontal and vertical axes denote first and second GDA coordinates. Five classes of data points corresponding to decreasing erosion risk – appearing from left (very high risk) to right (very low risk) – are marked differently. Generally, the topology of the subgroups is preserved and the groups appear fairly separated and condensed.

The overall pattern of the point configuration in Fig. 3 is the same as in Fig. 2. From left to right we see groups of points corresponding to areas with very high (vH), high
(H), medium (Me), low (L), and very low (vL) erosion risk. Generally, the topology of the subgroups is preserved. Both variates contribute significantly to the differentiation of the risk classes. Unfortunately, the model when applied to the test set, yields projections appearing in quite different areas; thus it is not able to make the generalization.

5 Supervised SOM

Kohonen's self-organizing maps are a popular tool for visualization of multivariate data. The method was successfully applied to the Kefallinia erosion data [1]. The SOM method uses a general purpose methodology without accounting specially for the additional information on class membership of the data points. However, after constructing the map, we may indicate by so called 'hits', what is the distribution (location) of the different classes. Map with hits of the classes is shown in Fig. 4 below.

![Sample1 Hits of 5 classes](image)

Fig. 4. Ordinary self-organizing map SOM of size 19 x 11 constructed from the samp1 learning data set using the Matlab SOM Toolbox by Vesanto et al. [10]. The erosion risk classes are neatly separated, with single overlapping hexagons. The erosion risk is progressing from the north (low risk) to the south (high risk).

Similarly as Fig. 3, also Fig. 4 was obtained using the data set samp1. When constructing the map, the class membership information was not used. The map was
created and graphed using the Matlab SOM Toolbox [10]. The same toolbox contains also another procedure, called 'som_supervised' (SOM_s), and based on a proposal by Kohonen et al. [7], how to include during the process of training the information on class membership. The procedure was added to the Matlab SOM Toolbox by Juha Parhankangas, who keeps the copyright of that procedure [10].

We have applied the 'som_supervised' technique to our data with the hope that it will ameliorate the already good differentiation among the classes. The result was negative: we got even a deterioration of the display.

We considered the idea that perhaps we should normalize our data in a different way, say statistically, to have the data with mean=0 and variance=1. Result: We got even more mixed classes.

The quality of a SOM is usually measured by two indices: the quantization error $q_e$ and the topographical error $t_e$ [10]. They are:

<table>
<thead>
<tr>
<th></th>
<th>$q_e$</th>
<th>$t_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordinary map:</td>
<td>0.0362</td>
<td>0.0327</td>
</tr>
<tr>
<td>Supervised map:</td>
<td>0.0453</td>
<td>0.1544</td>
</tr>
<tr>
<td>Ordinary normalized:</td>
<td>0.2203</td>
<td>0.0246</td>
</tr>
<tr>
<td>Supervised normalized:</td>
<td>0.2247</td>
<td>0.0596</td>
</tr>
</tbody>
</table>

Thus our conclusion: the best SOM quality is attained for the ordinary SOM.

### 6 Concluding Remarks

We compared in detail three methods serving for visualization of multivariate data, whose intrinsic dimension - as evaluated by the correlation fractal dimension - equals 1.60. This justifies the mapping of the data to a plane. The data were subdivided into 5 erosion risk classes and we wanted the mapping algorithm to take into account the class membership.

From the 3 investigated methods, the first one uses classical canonical discriminant functions (CDA alias LDA), which provide linear projections. The other two applied methods were: Generalized Discriminant Analysis (GDA) based on Gaussian kernels, and the som supervised SOM (SOM_s), a variant of Kohonen’s self-organizing map. All the 3 considered methods yielded similar results. In all projections, the erosion risk subgroups appeared fairly separated, as it should be. The GDA, by a proper tuning of the parameter ‘sigma’, yielded the classes more and more condensed and separated, however without generalization to other samples.

All the 3 methods preserved roughly the topology of the data, although the GDA has twisted sometimes the planar representation of the high and very high erosion group. The SOM_s appeared worse than the ordinary SOM, both in som quality and in differentiation of the risk classes. This is to a certain degree justified, because the ordinary SOM is trained to be optimal in the som quality, which means to be optimal in achieving both small quantization error and small topographic error. A change in conditions of the training may cause a deviation from optimality.