

Face Recognition

Jens Fagertun

Kongens Lyngby 2005
Master Thesis IMM-Thesis-2005-74

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

Abstract

This thesis presents a comprehensive overview of the problem of facial recognition. A survey of available facial detection algorithms as well as implementation and tests of different feature extraction and dimensionality reduction methods and light normalization methods are presented.

A new feature extraction and identity matching algorithm, the Multiple Individual Discriminative Models (MIDM) algorithm, is proposed.

MIDM is in collaboration with AAM-API, a C++ open source implementation of Active Appearance Models (AAM), implemented into the “FaceRec” Delphi 7 application, a real time automatic facial recognition system. AAM is used for face detection and MIDM for face recognition.

Extensive testing of the MIDM algorithm is presented and its performance evaluated by the Lausanne protocol. The Lausanne protocol is a precise and widely accepted protocol for the testing of facial recognition algorithms. These test evaluations showed that the MIDM algorithm is superior to all other algorithms reported by the Lausanne protocol.

Finally, this thesis presents a description of 3D facial reconstruction from a single 2D image. This is done by using prior knowledge in form of a statistical shape model of faces in 3D.

Keywords: Face Recognition, Face Detection, Lausanne Protocol, 3D Face Reconstruction, Principal Component Analysis, Fisher Linear Discriminant Analysis, Locality Preserving Projections, Kernel Fisher Discriminant Analysis.

Resumé

Denne afhandling præsenterer et omfattende overblik over problemet ansigts genkendelse. En oversigt over de tilgængelige algoritmer til detektering af ansigter såvel som implementation og test af forskellige metoder til ekstraktion af egenskaber og dimensionsreduktion samt metoder til lysnormalisering præsenteres.

En ny algoritme til ekstraktion af egenskaber og matchning af identiteter (Multiple Individual Discriminative Models - MIDM) er blevet foreslået.

MIDM, sammen med AAM-API, en open-source C++ implementering af Active Appearance Models (AAM), er blevet implementeret som applikationen "FaceRec" i Delphi 7. Denne applikation er et automatisk system til ansigts genkendelse, der kører i sand tid. AAM er brugt til ansigts detektering og MIDM er brugt til ansigts genkendelse.

Udførlig testning af MIDM algoritmen er præsenteret og dens ydelse evalueret ved hjælp af Lausanne protokollen. Lausanne protokollen er en præcis og bredt accepteret protokol for test af ansigts genkendelses algoritmer. Disse test evalueringer viste at MIDM algoritmen er alle andre algoritmer rapporteret ved hjælp af Lausanne protokollen overlegen.

Endeligt, præsenterer denne afhandling en beskrivelse af 3D ansigts rekonstruktion fra et enkelt 2D billede. Dette er gjort ved at bruge *a priori* kendskab i form af en statistisk model for formen af ansigter i 3D.

Nøgleord: Ansigts Genkendelse, Ansigts Detektering, Lausanne Protokollen, 3D Ansigts Rekonstruktion, Principal Komponent Analyse, Fisher Linear Dis-

kriminant Analyse, Locality Preserving Projections, Kernel Fisher Diskriminant Analyse.

Preface

This thesis was prepared at the Section for Image Analysis, in the Department of Informatics and Mathematical Modelling, IMM, located at the Technical University of Denmark, DTU, as a partial fulfillment of the requirements for acquiring the degree Master of Science in Engineering, M.Sc.Eng.

The thesis deals with different aspects of face recognition using both the geometrical and photometrical information of facial images. The main focus will be on face recognition from 2D images, but 2D to 3D conversion of data will also be considered.

The thesis consists of this report, a technical report and two papers; one published in *Proceedings of the 14th Danish Conference on Pattern Recognition and Image Analysis* and one submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*, written during the period January to September 2005.

It is assumed that the reader has a basic knowledge in the areas of statistics and image analysis.

Lyngby, September 2005

Jens Fagertun
[email: jens@fagertun.dk]

Publication list for this thesis

- [20] Jens Fagertun, David Delgado Gomez, Bjarne K. Ersbøll and Rasmus Larsen. A face recognition algorithm based on multiple individual discriminative models. *Proceedings of the 14th Danish Conference on Pattern Recognition and Image Analysis*, 2005.
- [21] Jens Fagertun and Mikkel B. Stegmann. The IMM Frontal Face Database. Technical Report, Informatics and Mathematical Modelling, Technical University of Denmark, 2005.
- [27] David Delgado Gomez, Jens Fagertun and Bjarne K. Ersbøll. A face recognition algorithm based on multiple individual discriminative models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. To appear - Submitted in 2005 - ID TPAMI-0474-0905.

Acknowledgements

I would like to thank the following people for there support and assistance in my preparation of the work presented in this thesis:

First and foremost, I thank my supervisor Bjarne K. Ersbøll for his support throughout this thesis. It has been an exciting experience and great opportunity to work with face recognition, a very interesting area in image analysis and pattern recognition.

I thank my co-supervisor Mikkel B. Stegmann for his huge initial support and always having time to spare.

I thank my good friend David Delgado Gomez for all the productive conversations on different issues of face recognition, and for an excellent partnership during the writing of the two papers included in this thesis.

I thank Rasmus Larsen for his great patience when answering questions of a statistical nature.

I thank my office-mates Mads Fogtmann Hansen, Rasmus Engholm and Steen Lund Nielsen for productive conversations and spending time answering my questions, which has been a great help.

I thank Mette Christensen and Bo Langgaard Lind for proof-reading the manuscript of this thesis.

I thank Lars Kai Hansen since he encouraged me to write my thesis at the image

analysis section.

In general I thank the staff of the image analysis- and computer graphics section for providing a pleasant and inspiring atmosphere as well as for their participation in the construction of the IMM Frontal Face Database.

Finally, I thank David Delgado Gomez and the Computational Imaging Lab at the Department of Technology at Pompeu Fabra University, Barcelona for their partnership in the participation in the ICBA2006¹ Face Verification Contest in Hong Kong in January, 2006.

¹International Conference on Biometrics 2006.

Contents

Abstract	i
Resumé	iii
Preface	v
Publication list for this thesis	vii
Acknowledgements	ix
1 Introduction	1
1.1 Motivation and Objectives	3
1.2 Thesis Overview	4
1.3 Mathematical Notation	5
1.4 Nomenclature	6
1.5 Abbreviations	7

I	Face Recognition in General	9
2	History of Face Recognition	11
3	Face Recognition Systems	13
3.1	Face Recognition Tasks	13
3.1.1	Verification	14
3.1.2	Identification	14
3.1.3	Watch List	14
3.2	Face Recognition Vendor Test 2002	16
3.3	Discussion	20
4	The Process of Face Recognition	21
4.1	Face Detection	22
4.2	Preprocessing	22
4.3	Feature Extraction	23
4.4	Feature Matching	23
4.5	Thesis Perspective	23
5	Face Recognition Considerations	25
5.1	Variation in Facial Appearance	25
5.2	Face Analysis in an Image Space	26
5.2.1	Exploration of Facial Submanifolds	27
5.3	Dealing with Non-linear Manifolds	28

5.3.1	Technical Solutions	28
5.4	High Input Space and Small Sample Size	30
6	Available Data	33
6.1	Face Databases	34
6.1.1	AR	34
6.1.2	BioID	34
6.1.3	BANCA	35
6.1.4	IMM Face Database	35
6.1.5	IMM Frontal Face Database	35
6.1.6	PIE	36
6.1.7	XM2VTS	36
6.2	Data Sets Used in this Work	36
6.2.1	Data Set I	37
6.2.2	Data Set II	37
6.2.3	Data Set III	41
6.2.4	Data Set IV	41
II	Assessment	45
7	Face Detection: A Survey	47
7.1	Representative Work of Face Detection	49
7.2	Description of Selected Face Detection Methods	49

7.2.1	General Aspects of Face Detections Algorithms	50
7.2.2	Eigenfaces	50
7.2.3	Fisherfaces	51
7.2.4	Neural Networks	51
7.2.5	Active Appearance Models	52
8	Preprocessing of a Face Image	59
8.1	Light Correction	59
8.1.1	Histogram Equalization	59
8.1.2	Removal of Specific Light Sources based on 2D Face Models	60
8.2	Discussion	62
9	Face Feature Extraction: Dimensionality Reduction Methods	65
9.1	Principal Component Analysis	66
9.1.1	PCA Algorithm	66
9.1.2	Computational Issues of PCA	68
9.2	Fisher Linear Discriminant Analysis	69
9.2.1	FLDA in Face Recognition Problems	70
9.3	Locality Preserving Projections	70
9.3.1	LPP in Face Recognition Problems	73
9.4	Kernel Fisher Discriminant Analysis	73
9.4.1	Problems of KFDA	77
10	Experimental Results I	79

10.1	Illustration of the Feature Spaces	80
10.2	Face Identification Tests	85
10.2.1	50/50 Test	85
10.2.2	Ten-fold Cross-validation Test	86
10.3	Discussion	87
III	Development	89
11	Multiple Individual Discriminative Models	91
11.1	Algorithm Description	92
11.1.1	Creations of the Individual Models	92
11.1.2	Classification	96
11.2	Discussion	96
12	Reconstruction of 3D Face from a 2D Image	99
12.1	Algorithm Description	99
12.2	Discussion	103
13	Experimental Results II	105
13.1	Overview	105
13.1.1	Initial Evaluation Tests	105
13.1.2	Lausanne Performance Tests	106
13.2	Initial Evaluation Tests	106
13.2.1	Identification Test	106

13.2.2 The Important Image Regions 108

13.2.3 Verification Test 110

13.2.4 Robustness Test 112

13.3 Lausanne Performance Tests 113

13.3.1 Participants in the Face Verification Contest, 2003 115

13.3.2 Results 116

13.4 Discussion 118

IV Implementation 121

14 Implementation 123

14.1 Overview 123

14.2 FaceRec 124

14.2.1 FaceRec Requirements 125

14.2.2 AAM-API DLL 125

14.2.3 Make MIDM Model File 126

14.3 Matlab Functions 126

14.4 Pitfalls 127

14.4.1 Passing Arrays 127

14.4.2 The Matlab Eig Function 127

V Discussion	129
15 Future Work	131
15.1 Light Normalization	131
15.2 Face Detection	132
15.3 3D Facial Reconstruction	132
16 Discussion	133
16.1 Summary of Main Contributions	133
16.1.1 IMM Frontal Face Database	133
16.1.2 The MIDM Algorithm	134
16.1.3 A Delphi Implementation	134
16.1.4 Matlab Functions	134
16.2 Conclusion	135
A The IMM Frontal Face Database	143
B A face recognition algorithm based on MIDM	153
C A face recognition algorithm based on MIDM	163
D FaceRec Quick User Guide	173
E AAM-API DLL	177
F CD-ROM Contents	181

Introduction

Face recognition is a task so common to humans, that the individual does not even notice the extensive number of times it is performed every day. Although research in automated face recognition has been conducted since the 1960's, it has only recently caught the attention of the scientific community. Many face analysis and face modeling techniques have progressed significantly in the last decade [30]. However, the reliability of face recognition schemes still poses a great challenge to the scientific community.

Falsification of identity cards or intrusion of physical and virtual areas by cracking alphanumeric passwords appear frequently in the media. These problems of modern society have triggered a real necessity for reliable, user-friendly and widely acceptable control mechanisms for the identification and verification of the individual.

Biometrics, which is based on authentication on the intrinsic aspects of a specific human being, appears as a viable alternative to more traditional approaches (such as PIN codes or passwords). Among the oldest biometric techniques is fingerprint recognition. This technique was used in China as early as 700 AD for official certification of contracts. Later on, in the middle of the 19th century, it was used for identification of persons in Europe [31]. A currently developed biometric technique is iris recognition [17]. This technique is now used instead of passport identification for frequent flyers in some airports in United King-

dom, Canada and the Netherlands. As well as for access control of employees to restricted areas in Canadian airports and in the New Yorks JFK airport. These techniques are inconvenient due to the necessity of interaction with the individual who is to be identified or authenticated. Face recognition on the other hand can be a non-intrusive technique. This is one of the reasons why this technique has caught an increased interest from the scientific community in the recent decade.

Facial recognition holds several advantages over other biometric techniques. It is natural, non-intrusive and easy to use. In a study considering the compatibility of six biometric techniques (face, finger, hand, voice, eye, signature) with machine readable travel documents (MRTD) [32] facial features scored the highest percentage of compatibility, see Figure 1.1. In this study parameters like the enrollment, renewal, machine requirements and public perception were considered. However, facial features should not be considered the most reliable biometric.

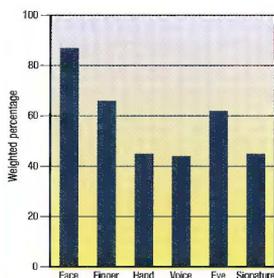


Figure 1.1: Comparison of machine readable travel documents (MRTD) compatibility with six biometric techniques; face, finger, hand, voice, eye, signature. Courtesy of Hietmeyer [32].

The increased interest automated face recognition systems have gained, from environments other than the scientific community is largely due to increasing public concerns for security, especially due to the many events of terror around the world after September 11th 2001.

However, automated facial recognition can be used in a lot of areas other than security oriented applications (access-control/verification systems, surveillance systems), such as computer entertainment and customized computer-human interaction. Customized computer-human interaction applications will in the near future be found in products such as cars, aids for disabled people, buildings, etc. The interest for automated facial recognition and the amount of applications will most likely increase even more in the future. This could be due to increased

penetration of technologies, such as digital cameras and the internet, and due to a larger demand for different security schemes.

Even though humans are experts in facial recognition is it not yet understood how this recognition is performed. For many years psychophysicists and neuroscientists have been researching whether face recognition is done holistically or by local feature analysis, i.e. is face recognition done by looking at the face as a whole or by looking at local facial features independently [6, 25]. It is however clear that humans are only capable of holding one face image in the mind at a given time. Figure 1.2 shows a classical illusion called “The Wife and the Mother-in-Law”, which was introduced into the psychological literature by Edwin G. Boring. What do you see? A witch or a young lady?



Figure 1.2: “The Wife and the Mother-in-Law” by Edwin G. Boring. What do you see? A witch or a young lady? Courtesy of Danial Chandler [8].

1.1 Motivation and Objectives

Face recognition has recently received a blooming attention and interest from the scientific community as well as from the general public. The interest from the general public is mostly due to the recent events of terror around the world, which has increased the demand for useful security systems. Facial recognition applications are far from limited to security systems as described above.

To construct these different applications, precise and robust automated facial

recognition methods and techniques are needed. However, these techniques and methods are currently not available or only available in highly complex, expensive setups.

The topic of this thesis is to help solving the difficult task of robust face recognition in a simple setup. Such a solution would be of great scientific importance and would be useful to the public in general.

The objectives of this thesis will be:

- To discuss and summarize the process of facial recognition.
- To look at currently available facial recognition techniques.
- To design and develop a robust facial recognition algorithm. The algorithm should be usable in a simple and easily adaptable setup. This implies a single camera setup, preferably a webcam, and no use of specialized equipment.

Besides these theoretical objectives a proof-of-concept implementation of the developed method will be carried out.

1.2 Thesis Overview

In the fulfilment with the objectives this thesis is naturally divided into five parts, where each part requires knowledge from the preceding parts.

Part I Face Recognition in General. Presents a summary of the history of face recognition. Discusses the different commercial face recognition systems, the general face recognition process and the different considerations regarding facial recognition.

Part II Assessment. Presents an assessment of the central tasks of face recognition identified in **Part I**, which include face detection, preprocessing of facial images and feature extracting.

Part III Development. Documents the design, development and testing of the Multiple Individual Discriminative Models face recognition algorithm. Furthermore, preliminary work in retrieval of depth information from one 2D image and a statistical shape model of 3D faces are presented.

Part IV Implementation. Documents the design and development of a face recognition system using the algorithm devised in **Part III**.

Part V Discussion. Presents a discussion of possible ideas to future work and conclude on the work done in this thesis.

1.3 Mathematical Notation

Throughout this thesis the following mathematical notations are used:

Scalar values are denoted with lower-case italic Latin or Greek letters:

$$x$$

Vectors, are denoted with lower-case, non-italic bold Latin or Greek letters. In this thesis only column vectors are used:

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T$$

Matrices are denoted with capital, non-italic bold Latin or Greek letters:

$$\mathbf{X} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Sets of objects such as scalars, vectors, images etc. are shown in vectors with curly braces:

$$\{a, b, c, d\}$$

Indexing into a matrix is displayed, as row-column subscript of either scalars or vectors:

$$\mathbf{M}_{xy} = \mathbf{M}_{\mathbf{x}}, \mathbf{x} = [x, y]$$

The mean vector of a specific data set, is denoted with lower-case, non-italic bold Latin or Greek letters with a bar:

$$\bar{\mathbf{x}}$$

1.4 Nomenclature

Landmarks set is a set of x and y coordinates that describes features (here facial features) like eyes, ears, noses, and mouth corners.

Geometric information is the distinct information of an object's shape, usually extracted by annotating the object with landmarks.

Photometric information is the distinct information of the image, i.e. the pixel intensities of the image.

Shape is according to Kendall [33] all the geometrical information that remains when location, scale and rotational effects are filtered out from an object.

Variables used throughout this thesis are listed below:

\mathbf{x}_i	A sample vector in the input space.
\mathbf{y}_i	A sample vector in the output space.
Φ	An eigenvector matrix.
ϕ_i	The i^{th} eigenvector.
Λ	A diagonal matrix of eigenvalues.
λ_i	The eigenvalue corresponding to the i^{th} eigenvector.
Σ	A covariance matrix.
\mathbf{S}_B	The between-class matrix, of Fisher Linear Discriminant Analysis.
\mathbf{S}_W	The within-class matrix, of Fisher Linear Discriminant Analysis.
\mathbf{S}	The adjacency graph, of Locality Preserving Projections.
Ψ	A non-linear mapping from an input space to a high dimensional implicit output space.
K	A Mercer kernel function.
\mathbf{I}	The identity matrix.

1.5 Abbreviations

A list of the abbreviations used in thesis can be found below:

PCA	Principal Component Analysis.
FLDA	Fisher Linear Discriminant Analysis.
LPP	Locality Preserving Projections.
KFDA	Kernel Fisher Discriminant Analysis.
MIDM	Multiple Individual Discriminative Models.
HE	Histogram Equalization.
FAR	False Acceptance Rate.
FRR	False Rejection Rate.
EER	Equal Error Rate.
TER	Total Error Rate.
CIR	Correct Identification Rate.
FIR	False Identification Rate.
ROC	Receiver Operating Characteristic (curve).
AAM	Active Appearance Model.
ASM	Active Shape Model.
PDM	Point Distribution Model.

Part I

Face Recognition in General

History of Face Recognition

The most intuitive way to carry out face recognition is to look at the major features of the face and compare these to the same features on other faces. Some of the earliest studies on face recognition were done by Darwin [15] and Galton [24]. Darwin's work includes analysis of the different facial expressions due to different emotional states, where as Galton studied facial profiles. However, the first real attempts to develop semi-automated facial recognition systems began in the late 1960's and early 1970's, and were based on geometrical information. Here, landmarks were placed on photographs locating the major facial features, such as eyes, ears, noses, and mouth corners. Relative distances and angles were computed from these landmarks to a common reference point and compared to reference data. In Goldstein *et al.* [26] (1971) a system is created of 21 subjective markers, such as hair color and lip thickness. These markers proved very hard to automate due to the subjective nature of many of the measurements still made completely by hand.

A more consistent approach to do facial recognition was done by Fischler *et al.* [23] (1973) and later by Yuille *et al.* [61] (1992). This approach measured the facial features using templates of single facial features and mapped these onto a global template.

In summary, most of the developed techniques during the first stages of facial recognition focused on the automatic detection of individual facial features. The

greatest advantages of these geometrical feature-based methods are the insensitivity to illumination and the intuitive understanding of the extracted features. However, even today facial feature detection and measurement techniques are not reliable enough for the geometric feature-based recognition of a face and geometric properties alone are inadequate for face recognition [12, 37].

Due to this drawback of geometric feature-based recognition, the technique has gradually been abandoned and an effort has been made in researching holistic color-based techniques, which has provided better results. Holistic color-based techniques align a set of different faces to obtain a correspondence between pixels intensities, a nearest neighbor classifier [16] can be used to classify new faces when the new image is first aligned to the set of already aligned images. By the appearance of the Eigenfaces technique [55], a statistical learning approach, this coarse method was notably enhanced. Instead of directly comparing the pixel intensities of the different facial images, the dimension of the input intensities were first reduced by a Principal Component Analysis (PCA) in the Eigenface technique. Eigenfaces is a basis component of many of the image based facial recognition schemes used today. One of the current techniques is Fisherfaces. This technique is widely used and referred [4, 9]. It combines the Eigenfaces with Fisher Linear Discriminant Analysis (FLDA) to obtain a better separation of the individual faces. In Fisherfaces, the dimension of the input intensity vectors is reduced by PCA and then FLDA is applied to obtain an optimal projection for separation of the faces from different persons. PCA and FLDA will be described in Chapter 9.

After development of the Fisherface technique, many related techniques have been proposed. These new techniques aim at providing an even better projection for separation of the faces from different persons. They try to strengthen the robustness in coping with differences in illumination or image pose. Techniques like Kernel Fisherfaces [59], Laplacianfaces [30] or discriminative common vectors [7] can be found among these approaches. The techniques behind Eigenfaces, Fisherfaces, Laplacianfaces and Kernel Fisherfaces will be discussed further later in this thesis.

CHAPTER 3

Face Recognition Systems

This chapter deals with the tasks of face recognition and how to report performance. The performance of some of the best commercial face recognition systems is included as well.

3.1 Face Recognition Tasks

The three primary face recognition tasks are:

- Verification (authentication) - Am I who I say I am? (one to one search)
- Identification (recognition) - Who am I? (one to many search)
- Watch list - Are you looking for me? (one to few search)

Different schemes are to be applied to test the three tasks described above. Which scheme to use depends on the nature of the application.

3.1.1 Verification

The verification task is aimed at applications requiring user interaction in the form of a identity claim, i.e. access applications.

The verification test is conducted by dividing persons into two groups:

- **Clients**, people trying to gain access using their own identity.
- **Imposters**, people trying to gain access using a false identity, i.e. an identity known to the system but not belonging to them.

The percentage of imposters gaining access is reported as the False Acceptance Rate (FAR) and an the percentage of client rejected access is reported as the False Rejection Rate (FRR) for a given threshold. An illustration of this is displayed in Figure 3.1.

3.1.2 Identification

The identification task is mostly aimed at applications not requiring user interaction, i.e. surveillance applications.

The identification test works from the assumption that all faces in the test are of known persons. The percentage of correct identifications is then reported as the Correct Identification Rate (CIR) or the percentage of false identifications is reported as the False Identification Rate (FIR).

3.1.3 Watch List

The watch list task is a generalization of the identification task which includes unknown people.

The watch list test is like the identification test reported in CIR or FIR, but can have FAR and FRR associated with it to describe the sensitivity of the watch list, meaning how often is an unknown classified as a person in the watch list (FAR).

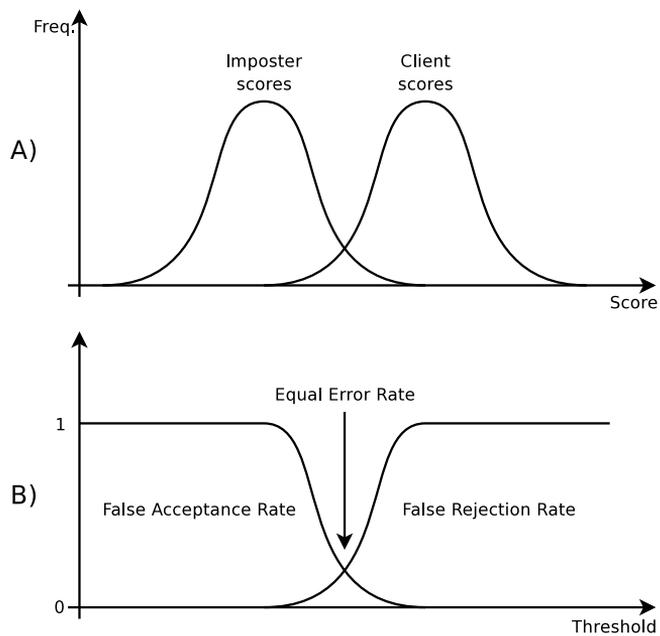


Figure 3.1: Relation of False Acceptance Rate (FAR), False Rejection Rate (FRR) with the distribution of clients, imposters in a verification scheme. A) Shows the imposters and client populations in terms of the score (high score meaning high likelihood of belonging to the client population). B) The associated FAR and FRR, the Equal Error Rate (EER) is where the FAR and FRR curve meets and gives the threshold value for the best separability of the imposter and client classes.

3.2 Face Recognition Vendor Test 2002

In 2002 the Face Recognition Vendor Test 2002 [45] tested some of the best commercial face recognition systems for their performance in the three primary face recognition tasks described in Section 3.1. This test used 121589 facial images of a group of 37437 different people. The different systems participating in the test are listed in Table 3.1. The evaluation was performed in reasonable controlled indoor lighting conditions¹.

Company	Web site
AcSys Biometrics Corp	http://www.acsysbiometricscorp.com
C-VIS GmbH	http://www.c-vis.com
Cognitec Systems GmbH	http://www.cognitec-systems.com
Dream Mirh Co., Ltd	http://www.dreammirh.com
Eyematic Interfaces Inc.	http://www.eyematic.com
Iconquest	http://www.iconquesttech.com
Identix	http://www.identix.com
Imagis Technologies Inc.	http://www.imagistechnologies.com
Viisage Technology	http://www.viisage.com
VisionSphere Technologies Inc.	http://www.visionspheretech.com

Table 3.1: Participants in the Face Recognition Vendor Test 2002.

¹Face recognition tests performed outside with unpredictable lighting conditions show a drastic drop in performance compared with indoor experiments [45].

The systems providing the best results in the vendor test show the characteristics listed in Table 3.2.

Tasks	CIR	FRR	FAR
Identification	73%		
Verification		10%	1%
Watch list	56% to 77% ²		1%

Table 3.2: The characteristics of the highest performing systems in the Face Recognition Vendor Test 2002. The highest performing system for the identification task and the watch list task was Cognitec. Cognitec and Identix was both the highest performing system for the verification task.

Selected conclusions from the Face Recognition Vendor Test 2002 are:

- The identification task yields better results for smaller databases, than larger ones. The identification task gave a higher score the smaller database used. Identification performance showed a linear decrease with respect to the logarithm of the size of the database. For every doubling of the size of the database performance decreased by 2% to 3%. See Figure 3.2.
- The face recognition systems showed a tendency to more easily identify older than younger people. The three best performing systems showed an average increase of performance by approximately 5% for every ten years increase of age of the test population. See Figure 3.3.
- The more time that elapses from the training of the system to the presentation of a new “up-to-date” image of a person the more recognition performance is decreased. For the three best performing systems there were an average decrease of approximately 5% per year. See Figure 3.4.

²56% and 77% corresponds to the use of watch lists of 3000 and 25 persons, respectively.

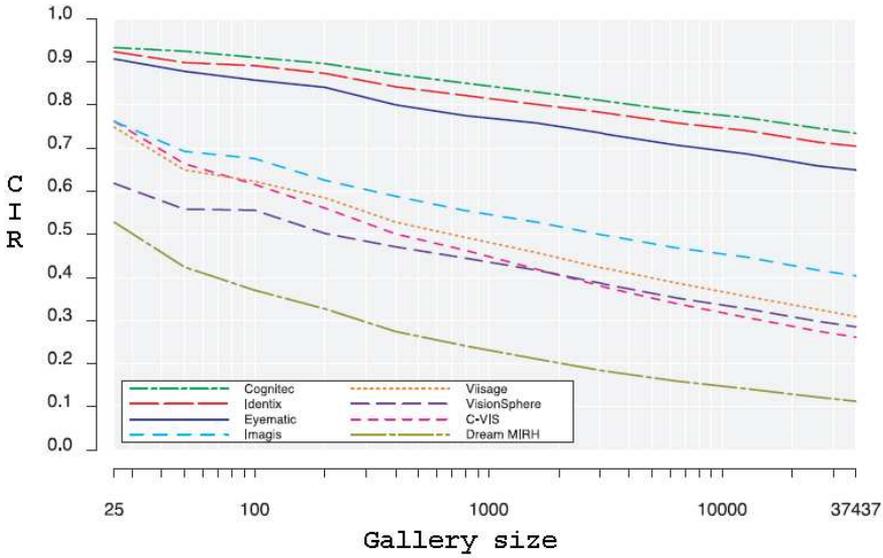


Figure 3.2: The Correct Identification Rates (CIR) plotted as a function of gallery size. Color of curves indicate the different vendors used in the test. Courtesy of Phillips *et al.* [45].

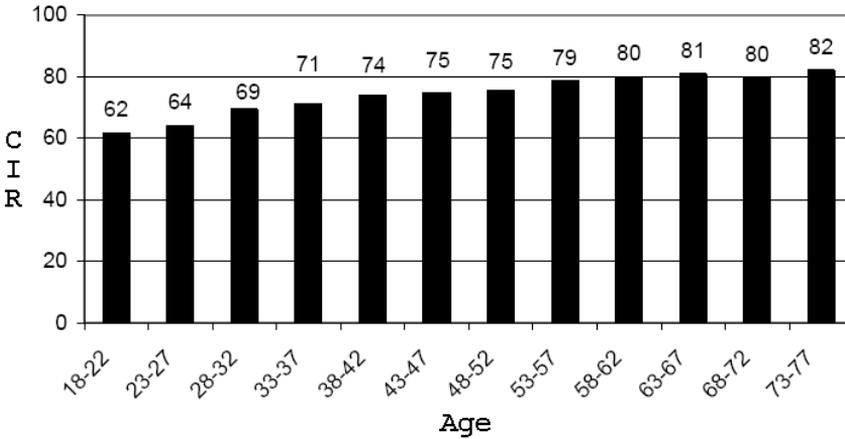


Figure 3.3: The average Correct Identification Rates (CIR) of the three highest performing systems (Cognitec, Identix and Eyematic), broken into age intervals. Courtesy of Phillips *et al.* [45].

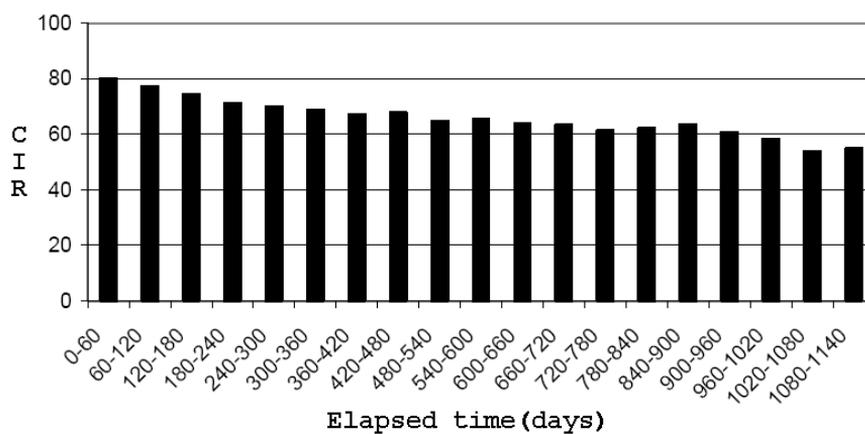


Figure 3.4: The average Correct Identification Rates (CIR) of the three highest performing systems (Cognitec, Identix and Eyematic), divided into intervals of elapsed time from the time of the systems construction to the time a new image is introduced to the systems. Courtesy of Phillips *et al.* [45].

3.3 Discussion

Interestingly, the results from the Face Recognition Vendor Test 2002 indicate a higher identification performance of older people compared to younger. In addition, the results indicate that it gets harder to identify people as time elapses, which is not surprising since the human face continually changes over time. The results of the Face Recognition Vendor Test 2002, reported in Table 3.2, are hard to interpret and compare to other tests, since change in the test protocol or test data will yield different results. However, these results provide an indication of the performance of commercial face recognition systems.

CHAPTER 4

The Process of Face Recognition

Facial recognition is a visual pattern recognition task. The three-dimensional human face, which is subject to varying illumination, pose, expression etc. has to be recognized. This recognition can be performed on a variety of input data sources such as:

- A single 2D image.
- Stereo 2D images (two or more 2D images).
- 3D laser scans.

Also, soon Time Of Flight (TOF) 3D cameras will be accurate enough to be used as well. The dimensionality of these sources can be increased by one by the inclusion of a time dimension. A still image with a time dimension is a video sequence. The advantage is that the identification of a person can be determined more precisely from a video sequence than from a picture since the identity of a person can not change from two frames taken in sequence from a video sequence.

This thesis is constrained to face recognition from single 2D images, even when tracking of faces is done in video sequences. However, Chapter 12 deals with

3D reconstruction of faces from one or more 2D images using statistical models of 3D laser scans.

Facial recognition systems usually consist of four steps, as shown in Figure 4.1; face detection (localization), face preprocessing (face alignment/normalization, light correction and etc.), feature extraction and feature matching. These steps are described in the following sections.

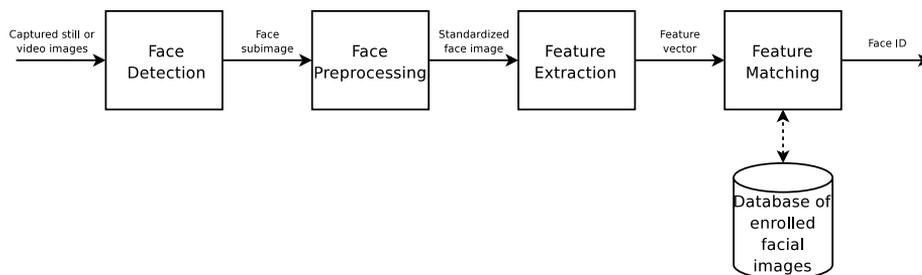


Figure 4.1: The four general steps in facial recognition.

4.1 Face Detection

The aim of face detection is localization of the face in a image. In the case of video input, it can be an advantage to track the face in between multiple frames, to reduce computational time and preserve the identity of a face (person) between frames. Methods used for face detection includes: Shape templates, Neural networks and Active Appearance Models (AAM).

4.2 Preprocessing

The aim of the face preprocessing step is to normalize the coarse face detection, so that a robust feature extraction can be achieved. Depending of the application, face preprocessing includes: Alignment (translation, rotation, scaling) and light normalization/correlation.

4.3 Feature Extraction

The aim of feature extraction is to extract a compact set of interpersonal discriminating geometrical or/and photometrical features of the face. Methods for feature extraction include: PCA, FLDA and Locality Preserving Projections (LPP).

4.4 Feature Matching

Feature matching is the actual recognition process. The feature vector obtained from the feature extraction is matched to classes (persons) of facial images already enrolled in a database. The matching algorithms vary from the fairly obvious Nearest Neighbor to advanced schemes like Neural Networks.

4.5 Thesis Perspective

This thesis will cover all four general areas in face recognition, though the primary focus is on feature extraction and feature matching.

A survey of face detection algorithms is presented in Chapter 7. Preprocessing of facial images is discussed in Chapter 8. A more in-depth description of feature extraction methods is presented in Chapter 9. The performance of these feature extraction methods is presented in Chapter 10, where the Nearest Neighbor algorithm will be used for feature matching. A new face recognition algorithm is developed in Chapter 11.

Face Recognition Considerations

In this chapter general considerations of the process of face recognition are discussed. These are:

- The variation of facial appearance of different individuals, which can be very small.
- The non-linear manifold on which face images reside.
- The problem of having a high-dimensional input space and only a small number of samples.

The scope of this thesis is further defined with the respect to these considerations.

5.1 Variation in Facial Appearance

A facial image is subject to various factors like facial pose, illumination and facial expression as well as lens aperture, exposure time and lens aberrations of

the camera. Due to these factors large variations of facial images of the same person can occur. On the other hand, sometimes small interpersonal variations occur. Here the extreme is identical twins, as can be seen in Figure 5.1. Different constraints in the process of acquiring images can be used to filter out some of these factors, as well as use of preprocessing methods.

In a situation where the variation among images obtained from the same person is larger than the variation among images of two individuals persons more comprehensive data than 2D images must be acquired to do computer based facial recognition. Here, accurate laser scans or infrared images (showing the blood vessel distribution in the face) can be used. These methods are out of the scope of this thesis and will not be discussed further. This thesis is mainly concerned with 2D frontal face images.



Figure 5.1: Small interpersonal variations illustrated by identical twins.
Courtesy of www.digitalwilly.com.

5.2 Face Analysis in an Image Space

When looking at the photometric information of a face, face recognition mostly rely on analysis of a subspace, since faces in images reside in a submanifold of the image space. This can be illustrated by an image consisting of 32×32 pixels. This image contains a total of 1024 pixels, with the ability to display a long range of different scenerys. Using only an 8-bit gray scale per pixel this image can show a huge number of different configurations, exactly $256^{1024} = 2^{8192}$. As a comparison the world population is only about 2^{32} . It is clear that only a small fraction of these image configurations will display faces. As a result most of the

original image space representation is very redundant from a facial recognition point of view. It must therefore be possible to reduce the input image space to obtain a much smaller subspace, where the objective of the subspace is to remove noise and redundancy while preserving the discriminative information of the face.

However, the manifolds where faces reside seem to be highly non-linear and non-convex [5, 53]. The following experiment explores this phenomenon in an attempt to obtain a deeper understanding of the problem.

5.2.1 Exploration of Facial Submanifolds

The purpose of the experiment presented in this section is to visualize that the facial images reside in a submanifold which is highly non-linear and non-convex.

For this purpose ten similar facial images were obtained from three persons of the IMM Frontal Face Database¹, yielding a total of 30 images. All images were converted to grayscale, cropped to only contain the facial region and scaled to 100×100 pixels. Then 33 new images were produced from each of the original images by following manipulations:

- **Translation;** Translation of the original image was done along the x-axis using the set (in pixels):

$$\{-30, -24, -18, -12, -6, 0, 6, 12, 18, 24, 30\}$$

- **Rotation;** Rotation of the original image was done around the center of the image using the set (in degrees):

$$\{-10, -8, -6, -4, -2, 0, 2, 4, 6, 8, 10\}$$

- **Scaling;** Scaling of the original image was done using the set (in %):

$$\{70, 76, 82, 88, 94, 100, 106, 112, 118, 124, 130\}$$

These manipulations resulted in the production of $30 \times 33 = 990$ images. An example of 33 images produced from one original image is shown in Figure 5.2. A Principal Component Analysis was conducted on the original 30 images to produce a three-dimensional subspace spanned by the three largest principal components. Then all 990 images were mapped into this subspace. These

¹This data set is further described in Chapter 6.

mappings into this subspace can be seen in Figure 5.3, where the images derived from the same original image are connected for easier visual interpretation. These mappings intuitively suggest that the manifold in which the facial images reside is non-linear and non-convex. A similar but more comprehensive test is performed by Li *et al.* [37].

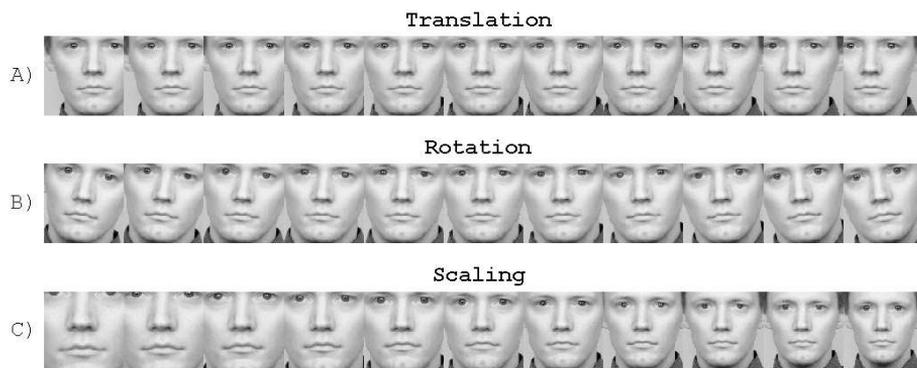


Figure 5.2: A sample of 33 facial images produced from one original image. The rows A, B and C are constructed by translation, rotation and scaling of the original image, respectively.

5.3 Dealing with Non-linear Manifolds

As described above is the face manifold highly non-linear and non-convex. The linear methods discussed later in Chapter 9 such as Principal Component Analysis (PCA) and Fisher Linear Discriminant Analysis (FLDA) are as a result only partly capable of preserving these non-linear variations.

5.3.1 Technical Solutions

To overcome the challenges of non-linear and non-convex face manifolds there are two general approaches:

- The first approach is to construct a feature subspace where the face manifolds become simpler, i.e. less non-linear and non-convex than the input space. This can be obtained by normalization of the face image both geometrically and photometrically to reduce variation. Followed by extraction

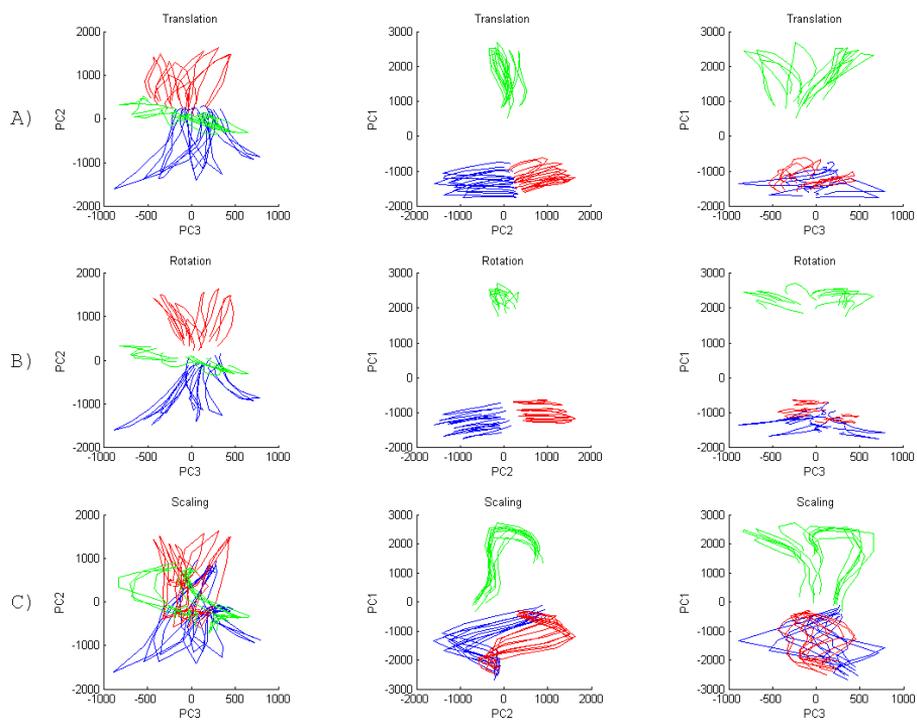


Figure 5.3: Results of the exploration of facial submanifolds. The 990 images derived from 30 original facial images are mapped into a three-dimensional space spanned by the three largest eigenvectors of the original images. The images derived from the original images are connected. The images of the three persons are plotted in different colors. The three sets of 30×11 images derived by translation, rotation and scaling are displayed in row A, B and C, respectively.

of features in the normalized image. For this purpose linear methods like PCA, FLDA or even non-linear methods as Kernel Fisher Discriminant Analysis (KFDA) can be used [1]. These methods will be described in Chapter 9.

- The second approach is to construct classification engines capable of solving the difficult non-linear classification problems of the image space. Methods like Neural Networks, Support Vector Machines etc. can be used for this purpose.

In addition the two approaches can be combined.

Work done using only the first approach to statistically understand and simplify the complex problem of facial recognition is pursued in this thesis.

5.4 High Input Space and Small Sample Size

Another problem associated with face recognition is the high input space of an image and the usually small sample size of an individual. An image consisting of 32×32 pixels resides in a 1024-dimensional space, where as the number of images of a specific person typically is much smaller. A small number of images of a specific person may not be sufficient to make an appropriate approximation of the manifold, which can cause a problem. An illustration of this problem is displayed in Figure 5.4. Currently, no known solution comes to mind for solving this problem. Other than capturing a sufficient number of samples to approximate the manifold in a satisfying way.

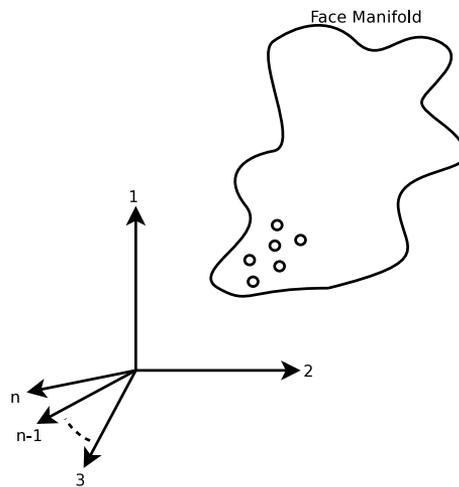


Figure 5.4: An illustration of the problem of not being capable of satisfactory approximating the manifold when only having a small number of samples. The samples are denoted by circles.

Available Data

This chapter presents a small survey of databases used for facial detection and recognition.

These databases include the IMM Frontal Face Database [21], which has been recorded and annotated with landmarks as a part of this thesis. The technical report made in conjunction with the IMM Frontal Face Database is found in Appendix A.

Finally, an in-depth description of the actual subsets of three databases used in this thesis is presented. The three databases used are:

- IMM Frontal Face Database: Used for initial testing in Chapter 10.
- The AR database: Used for a comprehensive test of the MIDM face recognition method (which is proposed in Chapter 11). The test results are shown in Chapter 13.
- The XM2VTS database: Used for evaluating the performance of the MIDM algorithm.

Work done using the XM2VTS database has been performed in collaboration

with Dr. David Delgado Gomez¹. The obtained results are to be used for the participation in the ICBA2006² Face Verification Contest in Hong Kong, January 2006.

6.1 Face Databases

In order to build/train and reliably test face recognition algorithms sizeable databases of face images are needed. Many face databases to be used for non-commercial purposes are available on the internet, either free of charge or for small fees.

These databases are recorded under various conditions and with various applications in mind. The following sections briefly describe some of the available databases which are widely known and used.

6.1.1 AR

The AR-database was recorded in 1998 at the Computer Vision Center in Barcelona. The database contains images of 116 people; 70 male and 56 female. Every person was recorded in two sessions each consisting of 13 images, resulting in a total of 3016 images. The two sessions were recorded two weeks apart. The 13 images of each session captured varying facial expressions, illuminations and occlusions. All images of the AR database are color images with a resolution of 768×576 pixels. Landmark annotations based on a 22-landmark scheme are available for some of the images of the AR database.

Link: "<http://rv11.ecn.purdue.edu/~aleix/aleix.face.DB.html>"

6.1.2 BioID

The BioID database was recorded in 2001. BioID contains 1521 images of 23 persons, about 66 images per person. The database was recorded during an unspecified number of sessions using a high variation of illumination, facial expression and background. The degree of variation was not controlled resulting

¹Post-doctoral at the Computational Imaging Lab, Department of Technology, Pompeu Fabra University, Barcelona.

²International Conference on Biometrics 2006.

in “real” life image occurrences. All images of the BioID database are recorded in grayscale with a resolution of 384×286 pixels. Landmark annotations based on a 20-landmark scheme are available.

Link: “<http://www.humanscan.de/support/downloads/facedb.php>”

6.1.3 BANCA

The BANCA multi database was collected as part of the European BANCA project. BANCA contains images, video and audio samples, though only the images are described here. BANCA contains images of 52 persons. Every person was recorded in 12 sessions each consisting of 10 images, resulting in a total of 6240 images. The sessions were recorded during a three months period. Three different image qualities were used to acquire the images, where each image quality was recorded during four sessions. All images are recorded in color with a resolution of 720×576 pixels.

Link: “<http://www.ee.surrey.ac.uk/banca/>”

6.1.4 IMM Face Database

The IMM Face Database was recorded in 2001 at the Department of Informatics and Mathematical Modelling - Technical University of Denmark. The database contains images of 40 people; 33 male and 7 female. It was recorded during one session and consists of 7 images per person resulting in a total of 240 images. The 7 images of each person were captured under varying facial expressions, camera view points and illuminations. Most of the images are recorded in color while the rest are recorded in grayscale, all with a resolution of 640×480 pixels. Landmark annotations based on a 58-landmark scheme are available.

Link: “http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3160”

6.1.5 IMM Frontal Face Database

The IMM Frontal Face Database was recorded in 2005 at the Department of Informatics and Mathematical Modelling - Technical University of Denmark. The database contains images of 12 people; all males. The database was recorded during one session and consists of 10 images of each person resulting in a total

of 120 images. The 10 images of each person were captured under varying facial expressions. All images are recorded in color with a resolution of 2560×1920 pixels. Landmark annotations based on a 73-landmark scheme are available.

Link: "http://www2.imm.dtu.dk/~aam/datasets/imm_frontal_face_db_high_res.zip"

6.1.6 PIE

The Pose, Illumination and Expression (PIE) database was recorded in 2000 at Carnegie Mellon University in Pittsburgh. The database contains images of 68 persons all recorded in one session. More than 600 images of each person were included in the database, resulting in a total of 41368 images. The images were captured under varying facial expressions, camera view points and illuminations. All images are recorded in color with a resolution of 640×468 pixels.

Link: "http://www.rh.cmu.edu/projects/project_418.html"

6.1.7 XM2VTS

The XM2VTS multi database was recorded at the University of Surrey. The database contains images, video and audio samples, though only the images are described here. XM2VTS contains images of 295 people. Every person was recorded during 4 sessions each consisting of four images per person, resulting in a total of 4720 images. The sessions were recorded during a four month period and captured both the frontal and the profiles of the face. All images are recorded in color with a resolution of 720×576 pixels.

Link: "<http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/>"

6.2 Data Sets Used in this Work

Three out of four data set used in this thesis are collected from face databases and consist of two parts: facial images and landmark annotations of the facial images. The last data set used in this thesis consists of 3D laser scans of faces. The next sections present the four data sets.

6.2.1 Data Set I

Data set I consists of the entire IMM Frontal Face Database [21]. In summary, this database contains 120 images of 12 persons (10 images a person). The 10 images of a person displays varying facial expressions, see Figure 6.1. The images have been annotated in a 73-landmark scheme, see Figure 6.2. A technical report of the construction of the database can be found in Appendix A.

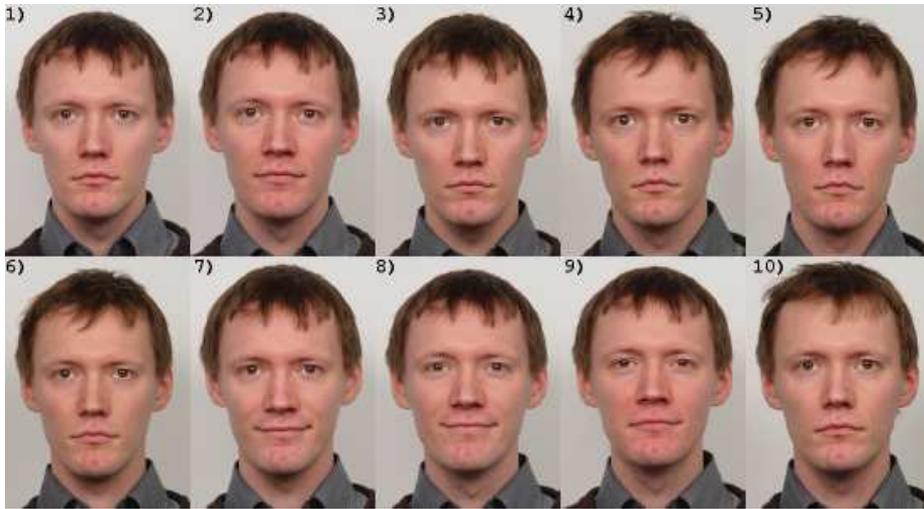


Figure 6.1: An example of ten images of one person from the IMM Frontal Face Database. The facial expressions of the images are: 1-6, neutral expression; 7-8, smiling (no teeth); 9-10, thinking.

6.2.2 Data Set II

Data set II consists of a subset of images from the AR database [41], where 50 persons (25 male and 25 female) were randomly selected. Fourteen images per person are included in **data set II**, which are obtained from the two recording sessions (seven images per person per session). The selected images were all images in the AR database without occlusions. **Data set II** is as a result composed of 700 images. Examples of the selected images of one male and one female from the two recording session are displayed in Figure 6.3.

Since no annotated landmarks were available for all the images of the AR-database, **data set II** required manually annotation using a 22-landmark scheme

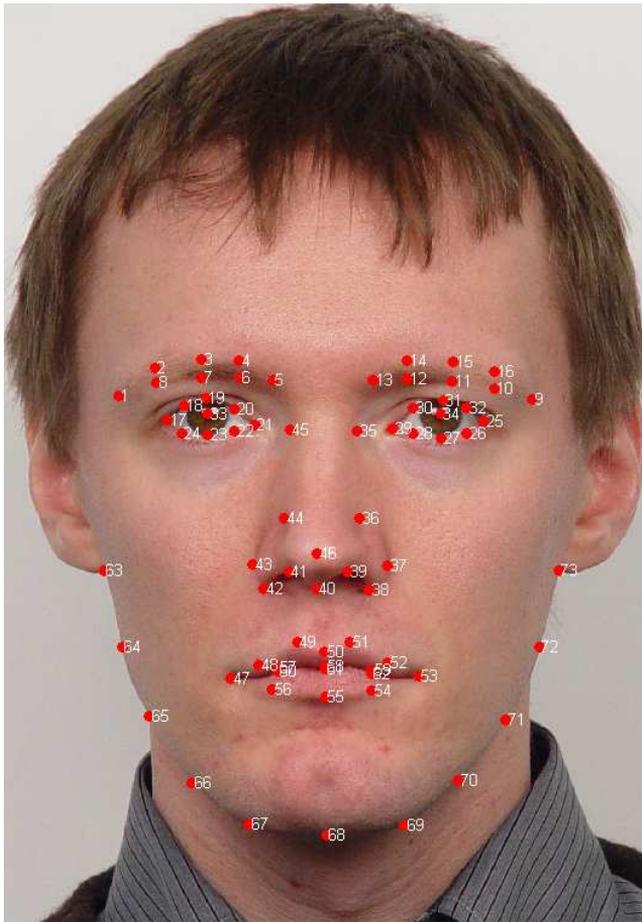


Figure 6.2: The 73-landmark annotation scheme used on the IMM Frontal Face Database.



Figure 6.3: Examples of 14 images of one female and one male obtained from the AR database. The rows of images (A, B) and (C, D) was captured during two different sessions. The columns display: 1, neutral expression; 2, smile; 3, anger; 4, scream; 5, left light on; 6, right light on; 7, both side lights on.

previously used by the Face and Gesture Recognition Working group³ (FGNET) to annotate parts of the AR database⁴. The 22-landmark scheme is displayed in Figure 6.4.

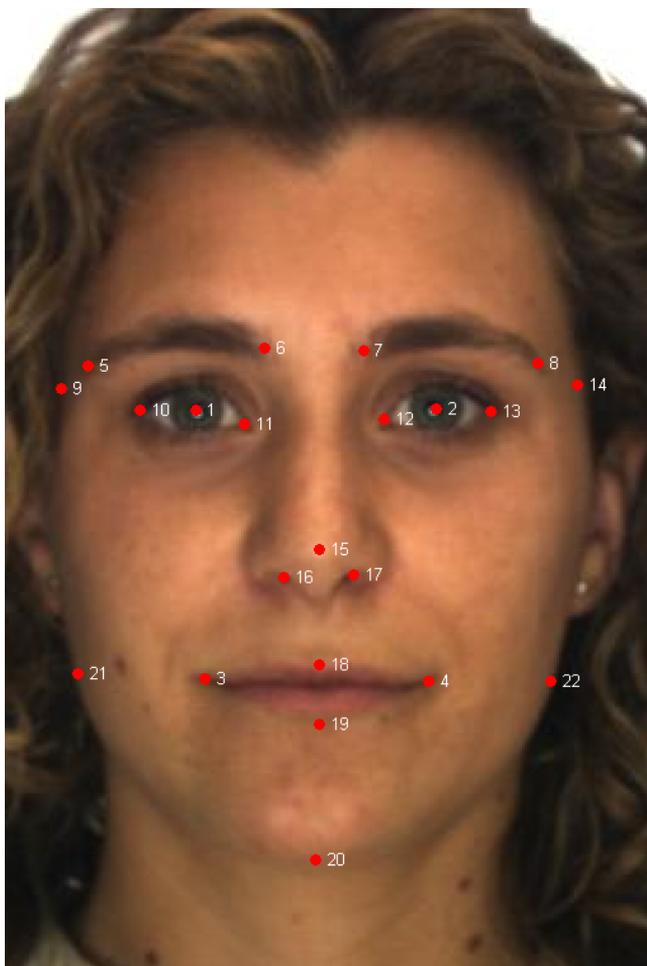


Figure 6.4: The 22-landmark annotation scheme used on the AR database.

³Link: "<http://www-prima.inrialpes.fr/FGnet/>".

⁴Of the 13 different image variations included in the AR database only 4 have been annotated by FGNET.

6.2.3 Data Set III

Data set III consists of all the frontal images from the XM2VTS database [43]. To summarize, 8 frontal images were captured of 295 individuals during 4 sessions, resulting in **data set III** consisting of a total of 2360 images. Examples of the selected images of one male and one female from the four recording session are displayed in Figure 6.5.

A 68-landmark annotation scheme is available for this data set, made in collaboration between the EU FP5 projects UFACE and FGNET. However, this thesis uses two non-public 64-landmark sets. The first set is obtained by manually annotation, where the second is obtained automatically by an optimized ASM [52]. Both landmark sets were created by the Computational Imaging Lab, Department of Technology, Pompeu Fabra University, Barcelona. The 64-landmark scheme is displayed in Figure 6.6.

6.2.4 Data Set IV

Data set IV consists of the entire 3D Face Database constructed by Karl Skoglund [49] at the Department of Informatics and Mathematical Modelling - Technical University of Denmark. This database includes 24 3D laser scans of 24 individuals (including one baby) and 24 texture images corresponding to the laser scans. Examples of five samples from **data set IV** are shown in Figure 6.7.



Figure 6.5: Examples of 8 images of one female and one male obtained from the XM2VTS database. All images are captured in a neutral expression.

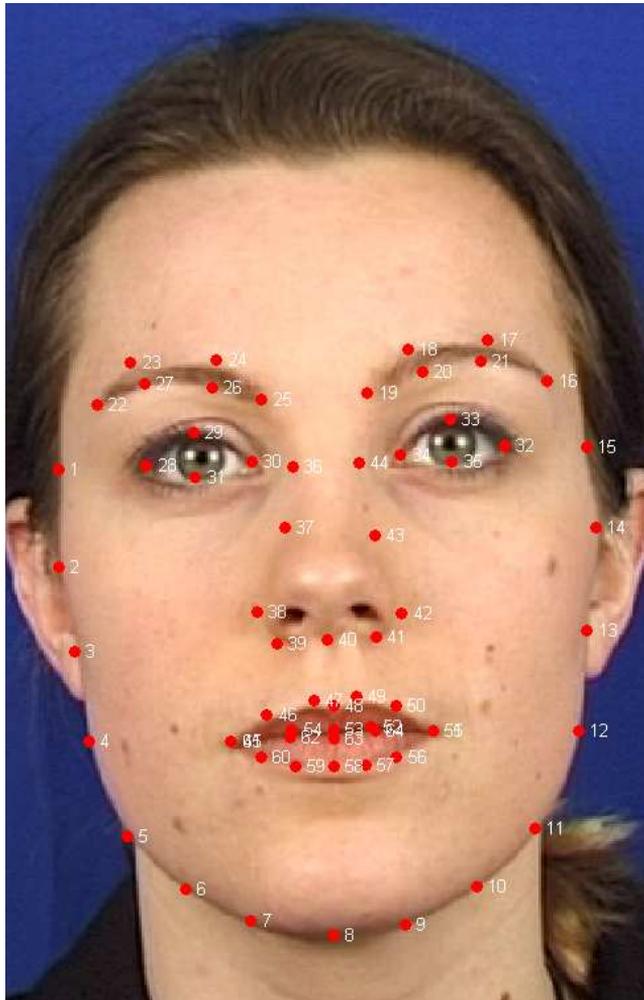


Figure 6.6: The 64-landmark annotation scheme used on the XM2VTS database.

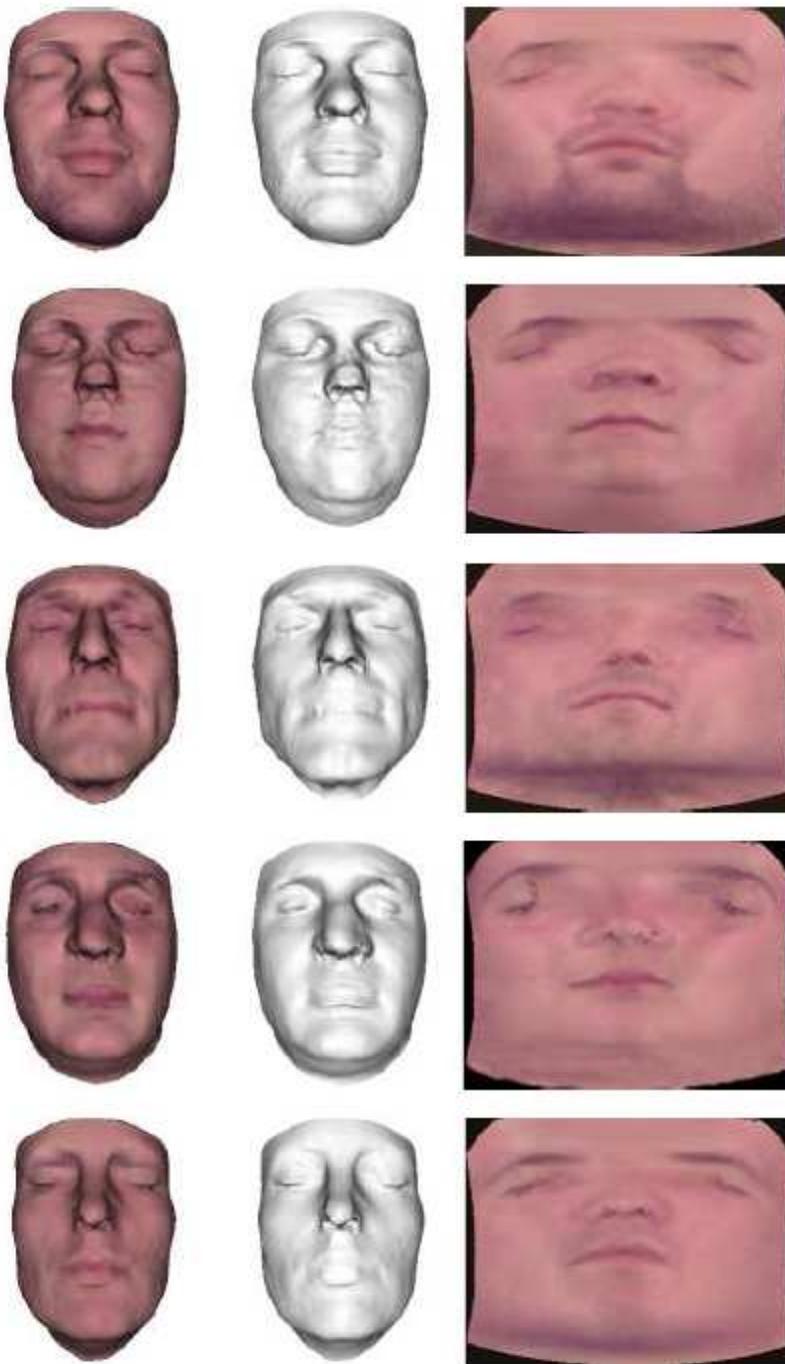


Figure 6.7: Five samples from 3D Face Database constructed in [49]. The 3D shape and texture, 3D shape and texture image is shown in the columns respectively.

Part II

Assessment

CHAPTER 7

Face Detection: A Survey

This chapter deals with the problem of face detection. Since the scope of this thesis is face recognition, this chapter will serve as an introduction to already developed algorithms for face detection.

As described earlier in Chapter 4, face detection is the necessary first step in a face recognition system. The purpose of face detection is to localize and extract the face region from the image background. However, since the human face is a highly dynamic object displaying large degree of variability in appearance, automatic face detection remains a difficult task.

The problem is complicated further by the continually changes over time of the following parameters:

- The three-dimensional position of the face.
- Removable features, such as spectacles and beards.
- Facial expression.
- Partial occlusion of the face, e.g. by hair, scarfs and sunglasses.
- Orientation of the face.

- Lighting conditions.

The following will distinguish between the two terms *face detection* and *face localization*.

Definition 7.1 *Face detection*, the process of detecting all faces (if any) in a given image.

Definition 7.2 *Face localization*, the process of localizing one face in a given image, i.e. the image is assumed to contain one, and only one face.

More than 150 methods for face detection have been developed, though only a small subset are addressed here. In Yang *et al.* [60] face detection methods are divided into four categories:

- **Knowledge-based methods:** The knowledge-based methods use a set of rules, that describe what to capture. The rules are constructed from the intuitive human knowledge of facial components and can be simple relations among facial features.
- **Feature invariant approaches:** The aim of feature invariant approaches is to search for structural features, which are invariant to changes in pose and lighting conditions.
- **Template matching methods:** Template matching methods constructs one or several templates (models) for describing facial features. The correlation between an input image and the constructed model(s) enables the method to discriminate over the case of face or non-face.
- **Appearance-based methods:** Appearance-based methods use statistical analysis and machine learning to extract the relevant features of a face to be able to discriminate between face and non-face images. The features are composed of both the geometrical information and the photometric information.

The knowledge-based methods and the feature invariant approaches are mainly used only for *face localization*, where as template matching methods and appearance-based methods can be used for *face detection* as well as *face localization*.

Approach	Representative Work
Knowledge-based	Multiresolution rule-based method [57]
Feature invariant - Facial Features - Texture - Skin Color - Multiple Features	Grouping of edges [36] Space Gray-Level Dependence matrix of face pattern [14] Mixture of Gaussian [58] Integration of skin color, size and shape [34]
Template matching - Predefined face templates - Deformable Templates	Shape templates [13] Active Shape Models [35]
Appearance-based method - Eigenfaces & Fisherfaces - Neural Network ¹ - Deformable Models	Eigenvector decomposition and clustering [54] Ensemble of neural networks and arbitration schemes [47] Active Appearance Models [10]

Table 7.1: Categorization of methods for face detection within a single image.

7.1 Representative Work of Face Detection

Representative methods of the four categories described above are summarized in Table 7.1 as reported in Yang *et al.* [60].

Only appearance-based methods are further described in this thesis since superior results seem to have been reported using these methods compared to the other three categories.

7.2 Description of Selected Face Detection Methods

In this section the methods of Eigenfaces, Fisherfaces, Neural Networks and Active Appearance Models are described, though with special emphasis on the Active Appearance Models. The Active Appearance Models show clear advantages for facial recognition purposes, which will be described and used later in this thesis.

¹Notice that neural networks are not restricted to appearance-based methods, but only neural networks working on photometrical information (texture) are considered here.

7.2.1 General Aspects of Face Detections Algorithms

Most face detection algorithms work by systemically analyzing subregions of an image. An example of how to extract these subregions could be, to capture a subimage of 20×20 pixels in the top left corner of the original image and continuing to capture subimages in a predefined grid. All these subimages are then evaluated using a face detection algorithm. Subsampling of the image in a pyramid fashion enables capture of different sizes face. This is illustrated in Figure 7.1.

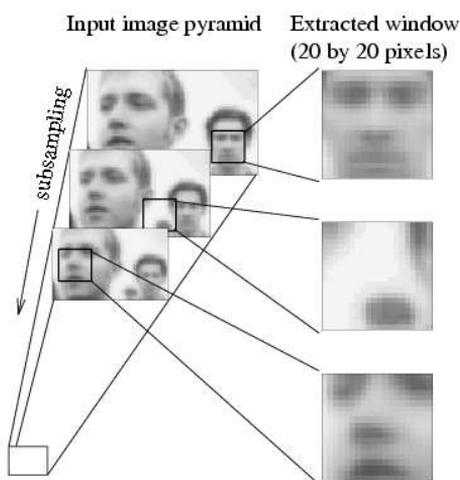


Figure 7.1: Illustration of the subsampling of an image in a pyramid fashion. Which enables the capture of different size of faces. Besides, rotated faces can be captured by rotating the subwindow. Courtesy of Rowley *et al.* [47].

7.2.2 Eigenfaces

The Eigenface method uses PCA to construct a set of Eigenface images. Examples of Eigenface images are displayed in Figure 7.2. These Eigenfaces, can be linear combined to reconstruct the images of the original training set. When introducing a new image an error (ξ) can be calculated from the best image reconstruction using the Eigenfaces to the new image. If the Eigenfaces are constructed from a large face database, the size of the error ξ can be used to determine whether or not a newly introduced image contains a face.

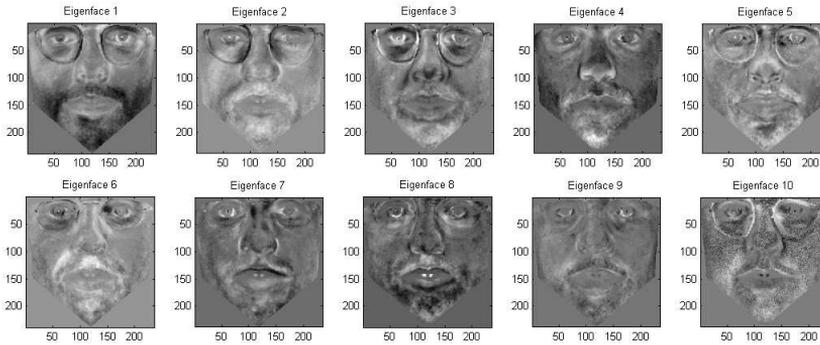


Figure 7.2: Example of 10 Eigenfaces. Notice that Eigenface no. 10 contains much noise and that the Eigenfaces are constructed from the shape free images described in Section 7.2.5.

Another more robust way is to look upon the subspace² provided by the eigenfaces, and cluster face images and non-face images in this subspace [54].

7.2.3 Fisherfaces

Much like Eigenfaces, Fisherfaces construct a subspace in which the algorithm can discriminate between facial and non-facial images. A more in-depth description of FLDA, which is used by Fisherfaces, can be found in Chapter 9.

7.2.4 Neural Networks

In a neural network approach features from an image are extracted and fed to a neural network. One huge drawback of neural networks is that they can be extensively tuned, in terms of deciding learning methods and on the number of layers, nodes, etc.

One of the most significant work in neural network face detection has been done by Rowley *et al.* [47, 48]. He used a neural network to classify images in a $[-1; 1]$ range, where -1 and 1 denotes a non-face image and a face image, respectively. Every image window of 20×20 pixels was divided into four 10×10 pixels, 16 5×5 pixels and six 20×5 pixels (overlapping) sub windows. A hidden node in the

²Principal Component Analysis can reduce the dimensionality of the data, described further in Chapter 9.

neural network was fed each of these sub windows, yielding a total of 26 hidden nodes. A diagram of the neural network design by Rowley *et al.* [47] is shown in Figure 7.3. The neural network can be improved by adding an extra neural network to determining the rotation of an image window. This will enable the system to capture faces not vertically aligned in the input image, see Figure 7.4.

7.2.5 Active Appearance Models

Active Appearance Models (AAM) are a generalization of the widely used Active Shape Models (ASM). Instead of only representing the information near edges, an AAM statistically models all texture and shape information inside the target model (here faces) boundary.

To build an AAM a training set has to be provided, which contains images and landmark annotations of facial features.

The first step in building an AAM is to align the landmarks using a Procrustes analysis [28], as displayed in Figure 7.5. Next the shape variation is modelled by a PCA, so that any shape can be approximated by

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \mathbf{b}_s, \quad (7.1)$$

where $\bar{\mathbf{s}}$ is the mean shape, Φ_s is a matrix containing the t_s most important eigenvectors and \mathbf{b}_s is a vector of length t_s , which contains a distinct set of parameters describing the actual shape. The number t_s of eigenvectors in Φ_s and the length of \mathbf{b}_s is chosen so that the model represents a user-defined proportion of the total variance in data. To obtain the proportion of p percent variance the value of t_s can be chosen by

$$\sum_{i=1}^{t_s} \lambda_i = > \frac{p}{100} \sum_{i=1}^n \lambda_i, \quad (7.2)$$

where λ_i is the eigenvalue corresponding to the i^{th} eigenvector and n is the total number of non-zero eigenvalues.

The texture variation is modelled by first removing shape information by warping all face images onto the mean shape. This is called the set of shape free images. Several methods can then be applied to eliminate global illumination

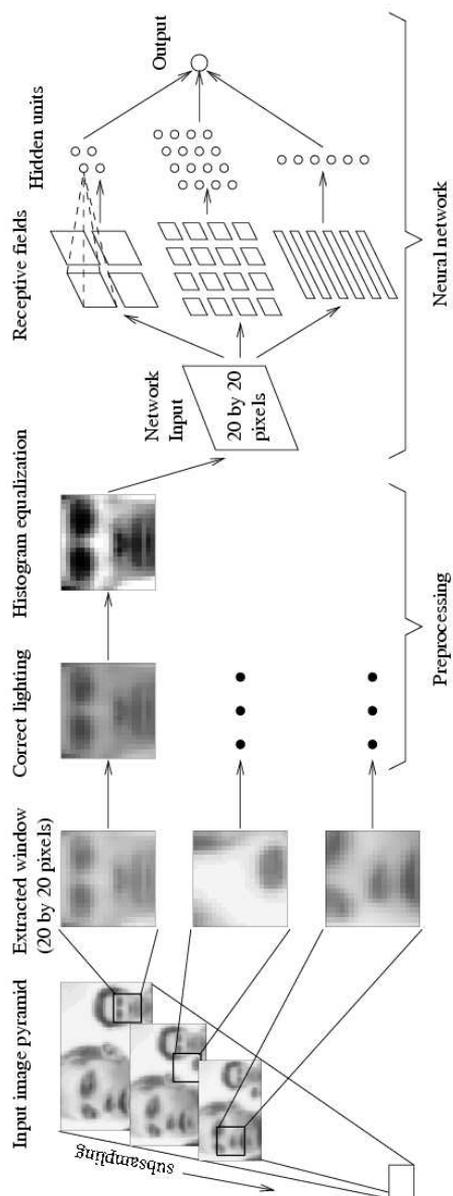


Figure 7.3: Diagram of the neural network developed by Rowley *et al.* [47].

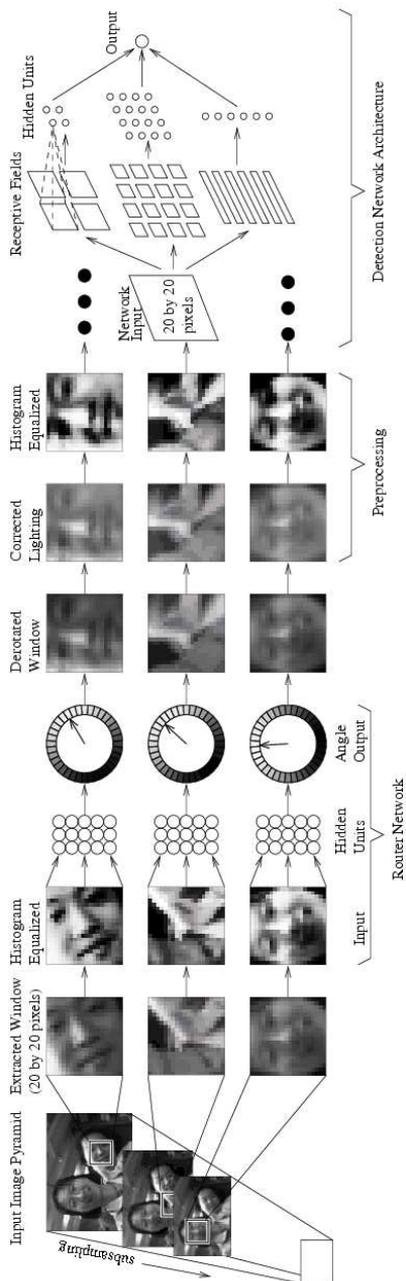


Figure 7.4: Diagram displaying an improved version of the neural network in Figure 7.3. Courtesy of Rowley *et al.* [48].

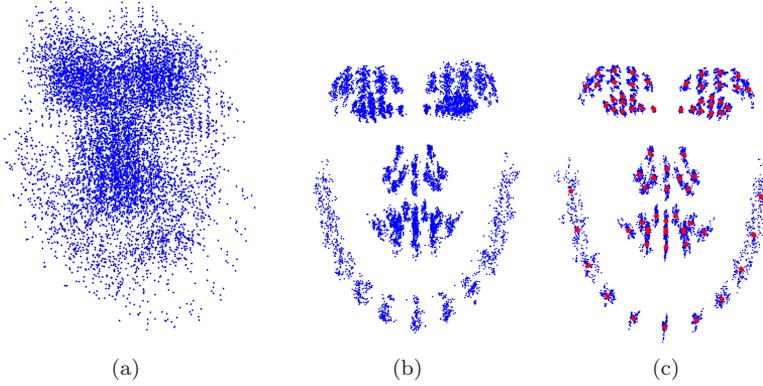


Figure 7.5: Full Procrustes analysis. (a) The original landmarks, (b) translation of the center of gravity (COG) into the mean shape COG, (c) result of full Procrustes analysis here the mean shape is plotted in red.

variation, see e.g. Cootes *et al.* [10]. Next, the texture variation can be modelled, like the shape by a PCA, so that any texture can be approximated by

$$\mathbf{t} = \bar{\mathbf{t}} + \Phi_t \mathbf{b}_t, \quad (7.3)$$

where $\bar{\mathbf{t}}$ is the mean texture, Φ_t is a matrix containing the t_t most important eigenvectors and \mathbf{b}_t is a vector of length t_t , which contains a distinct set of parameters describing the actual texture. t_t can be chosen, like t_s by Eq. 7.2.

The AAM is now built by concatenating shape and texture parameters

$$\mathbf{b} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \Phi_s^T (\mathbf{s} - \bar{\mathbf{s}}) \\ \Phi_t^T (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix}, \quad (7.4)$$

where \mathbf{W}_s is a diagonal matrix of weights between shape and texture. To remove the correlation between shape and texture a PCA is applied to obtain

$$\mathbf{b} = \Phi_c \mathbf{c}, \quad (7.5)$$

where \mathbf{c} is the AAM parameters. An arbitrary new shape and texture can be

generated by

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \mathbf{W}_s^{-1} \Phi_{c,s} \mathbf{c} \quad (7.6)$$

and

$$\mathbf{t} = \bar{\mathbf{t}} + \Phi_t \Phi_{c,t} \mathbf{c}, \quad (7.7)$$

where

$$\Phi_c = \begin{bmatrix} \Phi_{c,s} \\ \Phi_{c,t} \end{bmatrix}. \quad (7.8)$$

The process of placing the AAM mean shape and texture on a specific location in an image and search for a face near by this location, is shown in Figure 7.6. This process will not be described further here. For a more detailed description of AAM the paper Cootes *et al.* [10] or the master thesis by Mikkel Bille Stegmann [50] are recommended.

One advantage of the AAM (and ASM) algorithm compared to other face detection algorithms is that a localized face is described both by shape and texture.

Thus, a well defined shape of the face can be obtained by an AMM. This is an improvement from others face detection algorithms, where the result is a sub image containing a face without knowing exactly which pixels represent background and which represent the face. An AAM is also desirable for tracking in video sequences, assuming that changes are minimal from frame to frame. Due to these advantages an AAM is used as the face detection algorithm in this thesis, when automatic detection is required.

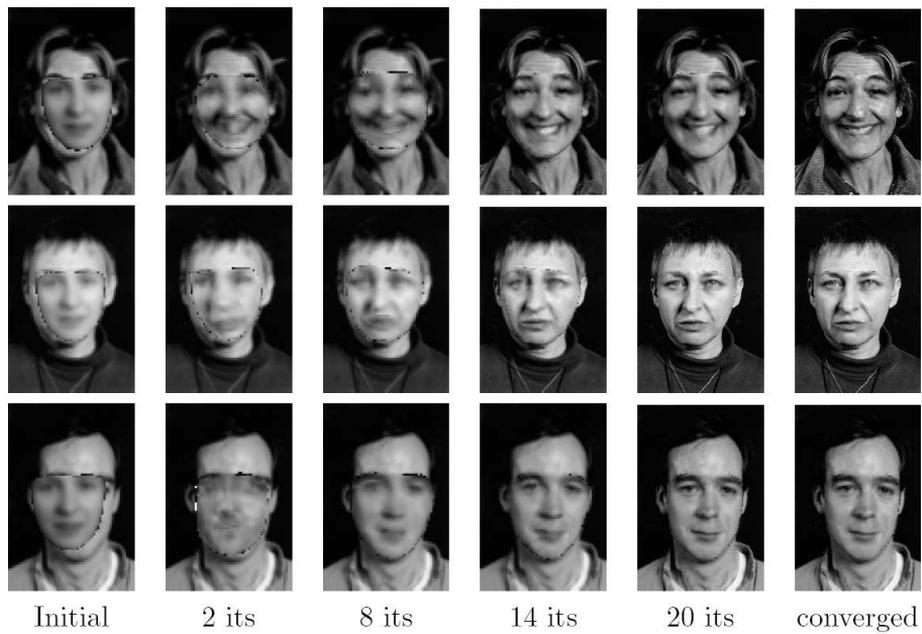


Figure 7.6: Face detection (approximations) obtained by AAM, when the model is initialized close to the face. The first column is the mean shape and texture of the AAM. The last column is the converged shape and texture of the AAM. Courtesy Cootes *et al.* [10].

Preprocessing of a Face Image

The face preprocessing step aims at normalizing, i.e. reducing the variation of images obtained during the face detection step. Using AAM in the process of face detection provides a well defined framework to retrieve the photometric information as a shape free image as well as the geometric information as a shape. Since this already has been described previously in this thesis only the subject of light correction will be described within this chapter.

8.1 Light Correction

As described in Section 3.2, unpredictable change in lighting conditions is a problem in facial recognition. Therefore, it is desirable to normalize the photometric information in terms of light correction to optimize the facial recognition. Here, two light correction methods are described.

8.1.1 Histogram Equalization

Histogram equalization (HE) can be used as a simple but very robust way to obtain light correction when applied to small regions such as faces. The aim of

HE is to maximize the contrast of an input image, resulting in a histogram of the output image which is as close to a uniform histogram as possible. However, this does not remove the effect of a strong light source but maximizes the entropy of an image, thus reducing the effect of differences in illumination within the same “setup” of light sources. By doing so, HE makes facial recognition a somehow simpler task. Two examples of HE of images can be seen in Figure 8.1. The algorithm of HE is straight forward and will not be explained here, an interested reader can obtain the algorithm in Finlayson *et al.* [22].

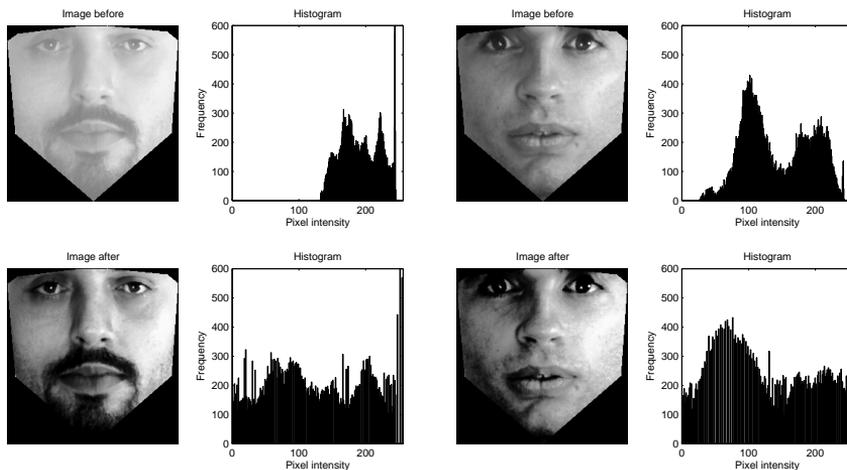


Figure 8.1: Examples of histogram equalization used upon two images to obtain standardized images with maximum entropy. Notice, only the facial region of an image is used in the histogram equalization.

8.1.2 Removal of Specific Light Sources based on 2D Face Models

The removal of specific light sources based on 2D face models [56] is another method to obtain light correlation of images. The method creates a pixelwise correspondence of images (as already described in Section 7.2.5, the AAM shape free image). By doing so, the effect of illumination upon each pixel $\mathbf{x} = \{x, y\}$ of an image can be expressed by the equation

$$\tilde{\mathbf{F}}_{\mathbf{x}} = a_{i,\mathbf{x}} \cdot \mathbf{F}_{\mathbf{x}} + b_{i,\mathbf{x}}, \quad (8.1)$$

where \mathbf{F} and $\tilde{\mathbf{F}}$ are the images of the same scene recoded at normal lighting condition (diffuse lighting) and upon the influence of a specific light source (illumination mode i), respectively. $a_{i,\mathbf{x}}$ is the multiplication compensation, and $b_{i,\mathbf{x}}$ is the additive compensation of the illumination mode i of pixel \mathbf{x} in the image $\tilde{\mathbf{F}}$.

Having n sets of images in the normal illumination and the mode i illumination, Eq. 8.1 can be rewritten as

$$\tilde{\mathbf{G}} = \mathbf{G} \begin{bmatrix} a_{i,\mathbf{x}} \\ b_{i,\mathbf{x}} \end{bmatrix}, \quad (8.2)$$

where

$$\tilde{\mathbf{G}} = \begin{bmatrix} \tilde{\mathbf{F}}_{1,\mathbf{x}} \\ \vdots \\ \tilde{\mathbf{F}}_{n,\mathbf{x}} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{F}_{1,\mathbf{x}} & 1 \\ \vdots & \vdots \\ \mathbf{F}_{n,\mathbf{x}} & 1 \end{bmatrix}. \quad (8.3)$$

If the n sets of images are of different persons, then the rows of \mathbf{G} and $\tilde{\mathbf{G}}$ are independent and the least-squares solution to $a_{i,\mathbf{x}}$ and $b_{i,\mathbf{x}}$ in Eq. 8.2 is

$$\begin{bmatrix} a_{i,\mathbf{x}} \\ b_{i,\mathbf{x}} \end{bmatrix} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \tilde{\mathbf{G}}. \quad (8.4)$$

Using Eq. 8.4 upon every pixel in the shape free image, the illumination compensation images \mathbf{A}_i and \mathbf{B}_i can be constructed. By doing so it is possible to reconstruct a face image in normal lighting conditions from a face image in lighting condition i by

$$\mathbf{F}_{\mathbf{x}} = \frac{\tilde{\mathbf{F}}_{\mathbf{x}} - b_{i,\mathbf{x}}}{a_{i,\mathbf{x}}}. \quad (8.5)$$

Different schemes can be used to identify the lighting condition of a specific face image, in Xie *et al.* [56] a FLDA is used.

Removal of two specific illumination conditions is displayed in Figure 8.2. This used the illumination compensation maps displayed in Figure 8.4. However, this

method sometimes creates artifacts in the faces. A close-up of the illumination corrected faces from Figure 8.2 can be seen in Figure 8.3 that displays this fact.

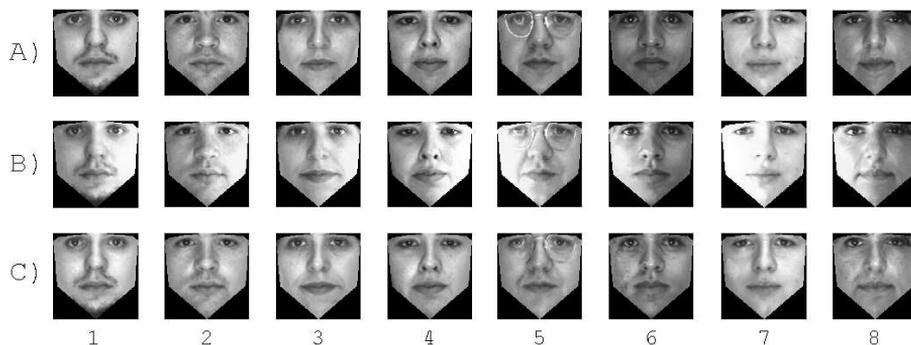


Figure 8.2: Removal of specific illumination conditions from facial images. A) shows the facial images in normal diffuse lighting. B) column 1-4 and 5-8 show facial images captured under right and left illumination, respectively. C) is the compensated images.

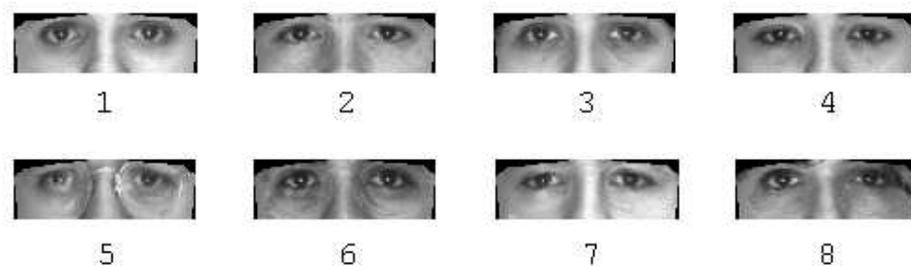


Figure 8.3: A close-up of the faces reconstructed in Figure 8.2. Notice that faces 1-4 are influenced only little by artifacts while faces 5-8 are influenced substantially by artifacts.

8.2 Discussion

It is clear that HE is a good and robust way of normalizing images. The more complex method of removing specific illumination conditions seems to yield impressive results, but has the drawback of sometimes imposing artifacts onto the images, as can be seen in Figure 8.3, where “shadows of spectacles” can be seen on persons not wearing spectacles. It was decided to only preprocess

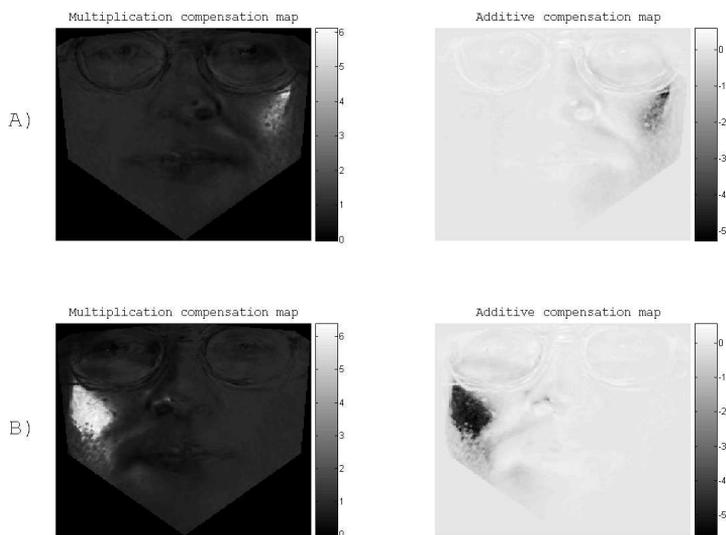


Figure 8.4: Illumination compensation maps used for removal of specific illumination conditions. Rows A) and B) display the illumination compensation maps for facial images captured under left and right illumination, respectively.

facial images with HE to ensure that the images are independent. No tests were performed to see how facial recognition performs under the influence of the artifacts introduced by the removal of specific light sources based on 2D face models. This will be saved for future work.

Face Feature Extraction: Dimensionality Reduction Methods

Table 9.1 lists the most promising dimensionality reduction methods (feature extraction methods) used for face recognition. Out of these Principal Component Analysis, Fisher Linear Discriminant Analysis, Kernel Fisher Linear Discriminant Analysis and Locality Preserving Projections will be described in the following.

Preserving	Technique	Method
Global Structure	Linear	Fisher Linear Discriminant Analysis
		Principal Component Analysis
	Non-linear	Kernel Fisher Linear Discriminant Analysis
		Kernel Principal Component Analysis
Local Structure	Linear	Locality Preserving Projections
	Non-linear	Isomap
		Laplacian Eigenmap

Table 9.1: Dimensionality reduction methods.

9.1 Principal Component Analysis

Principal Component Analysis (PCA), also known as Karhunen-Loève transformation, is a linear transformation which captures the variance of the input data. The coordinate system in which the data resides is rotated by PCA, so that the first-axis is parallel to the highest variance in the data (in a one-dimension projection). The remaining axes can be explained one at the time as being parallel to the highest variance of the data, while all axes are constrained to be orthogonal to all previous found axes. To summarize, the first-axis will contain highest variance, the second-axis contain the second highest variance, etc. An example in two dimensions is shown in Figure 9.1. PCA, which is an unsupervised method, is a powerful tool for data analysis, especially if data resides in a space higher than three dimensions, where graphical representations are hard. One of the main applications of PCA is dimension reduction, with little or no loss of data variation. This is used to remove redundancy and compress data.

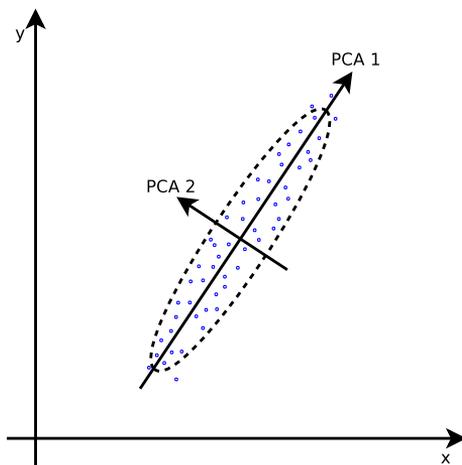


Figure 9.1: An example of PCA in two dimensions, showing the PCA axis that maximizes the variation in the first principal component: PCA 1.

9.1.1 PCA Algorithm

Different methods can be used to calculate the PCA basis vectors. Here eigenvalues and eigenvectors of the covariance matrix of the data are used. Considering

the data

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n], \quad (9.1)$$

where n is the amount of data samples, \mathbf{x}_i is the i th data sample of dimension d . First is the mean of \mathbf{X} subtracted from the data

$$\hat{\mathbf{X}} = [\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_2 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}}]. \quad (9.2)$$

The covariance matrix $\Sigma_{\hat{\mathbf{X}}}$ is calculated by

$$\Sigma_{\hat{\mathbf{X}}} = \frac{1}{n} \hat{\mathbf{X}} \hat{\mathbf{X}}^T. \quad (9.3)$$

The principal axes are now given by the eigenvectors $\Phi_{\hat{\mathbf{X}}}$ of the covariance matrix

$$\Sigma_{\hat{\mathbf{X}}} \Phi_{\hat{\mathbf{X}}} = \Phi_{\hat{\mathbf{X}}} \Lambda_{\hat{\mathbf{X}}}, \quad (9.4)$$

where

$$\Lambda_{\hat{\mathbf{X}}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_d \end{bmatrix} \quad (9.5)$$

is the diagonal matrix of eigenvalues corresponding to the eigenvectors of

$$\Phi_{\hat{\mathbf{X}}} = [\phi_1, \phi_2, \dots, \phi_d]. \quad (9.6)$$

The eigenvector corresponding to the highest eigenvalue represents the basis vector containing the most data variance, i.e. the first principal component.

The i th data sample, \mathbf{x}_i , can be transformed into the PCA space by

$$\mathbf{y}_i = \Phi_{\hat{\mathbf{X}}}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) = \Phi_{\hat{\mathbf{X}}}^T(\mathbf{x}_i - \bar{\mathbf{x}}). \quad (9.7)$$

Notice that an orthogonal matrix as $\Phi_{\hat{\mathbf{X}}}$ has the property $\Phi_{\hat{\mathbf{X}}}^{-1} = \Phi_{\hat{\mathbf{X}}}^T$. Data in the PCA space can be transformed back into the original space by

$$\mathbf{x}_i = \Phi_{\hat{\mathbf{X}}}\mathbf{y}_i + \bar{\mathbf{x}}. \quad (9.8)$$

If only a subset of the eigenvectors in $\Phi_{\hat{\mathbf{X}}}$ is selected, then this will result in data being projected into a PCA subspace. This can be very useful to reduce redundancy in the data, i.e. remove all eigenvectors equal to zero. The above method is described in greater detail in Ersbøll *et al.* [19].

9.1.2 Computational Issues of PCA

If one has n data samples in a d high-dimensional space where $n \ll d$. Then the computational time is quite large for retrieving eigenvectors and eigenvalues from the $d \times d$ covariance matrix. The time needed for eigenvector decomposition increases by the cube of the covariance matrix size [10]. However, it is possible to calculate the eigenvectors of the non-zero eigenvalues from a much smaller matrix with size $n \times n$, by use of

$$\Sigma_n = \frac{1}{n} \hat{\mathbf{X}}^T \hat{\mathbf{X}}, \quad (9.9)$$

where $\hat{\mathbf{X}}$ is calculated by Eq. 9.2. The non-zero eigenvalues of the matrices in Eq. 9.4 and Eq. 9.9 are equal

$$\Lambda_n = \Lambda_{\hat{\mathbf{X}}}. \quad (9.10)$$

The eigenvectors corresponding to non-zero eigenvalues can be expressed as

$$\hat{\Phi}_{\hat{\mathbf{X}}} = \hat{\mathbf{X}} \Phi_n. \quad (9.11)$$

Notice that these eigenvectors are not normalized. This can be proved by the Eckhart-Young Theorem [50].

9.2 Fisher Linear Discriminant Analysis

Fisher Linear Discriminant Analysis (FLDA), also known as Canonical Discriminant Analysis is like PCA, a linear transformation. Unlike PCA, FLDA is a supervised method, which implies that all training-data samples must be associated (manually) with a class. FLDA maximizes the between-class variance as well as minimizes the within-class variance. A graphic example of FLDA is shown in Figure 9.2.

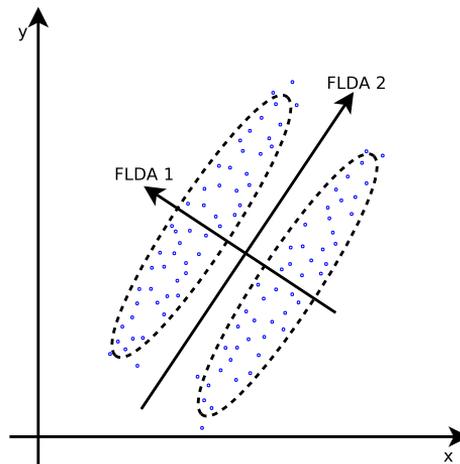


Figure 9.2: An example of FLDA in two dimensions, showing the FLDA axis that maximizes the separation between the classes and minimizes the variation inside the classes.

The objective function for FLDA is as follows

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}, \quad (9.12)$$

where the between-matrix is defined as

$$\mathbf{S}_B = \sum_{i=1}^c n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad (9.13)$$

and the within-matrix as

$$\mathbf{S}_W = \sum_{i=1}^c \sum_{j=1}^{n_i} (\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)^T, \quad (9.14)$$

where $\mathbf{x}_{i,j}$ is the j^{th} sample in class i , $\bar{\mathbf{x}}_i$ mean of class i , $\bar{\mathbf{x}}$ mean of all samples, c is number of classes and n_i is the number of samples in class i .

The optimal projection that maximizes the between-class variance and minimizes the within-class variance is given by the direction of the eigenvector associated to the maximum eigenvalue of $\mathbf{S}_W^{-1}\mathbf{S}_B$. Notice that the number of non-zero eigenvalues is at most number of classes minus one [4].

9.2.1 FLDA in Face Recognition Problems

In face recognition problems \mathbf{S}_W is nearly always singular. This is due to the fact that the rank of \mathbf{S}_W is at most $n - c$, where n (the number of training samples) usually is much smaller than the number of pixels in each image.

In order to overcome this problem a PCA is usually performed¹ on the images prior to FLDA, which removes redundancy and makes the data samples more compact. The within-matrix, \mathbf{S}_W , is made non-singular by only considering the f most important principal components from the PCA, where f is the number of non-zero eigenvalues of the within-matrix \mathbf{S}_W .

9.3 Locality Preserving Projections

The Locality Preserving Projections (LPP) algorithm has recently been developed [29]. When high-dimensional data lies on a low dimension manifold

¹Normally capturing between 95-99% of the variance.

embedded in the data space, then LPP approximate the eigenfunctions of the Laplace-Beltrami operator of the manifold. LPP aims at preserving the local structure of the data. This is unlike PCA and FLDA, which aims at preserving the global structure of the data.

LPP is unsupervised and performs a linear transformation. It models the manifold structure by constructing an adjacency graph, which is a graph expressing local nearness of the data. This is highly desirable for face recognition compared to non-linear local structure preserving methods in Table 9.1, since it is significantly less computationally expensive and more importantly it is defined in all points and not just in the training points as Isomaps and Laplacian Eigenmaps.

The objective function of LPP is

$$\min \sum_{ij} (y_i - y_j)^2 \mathbf{S}_{ij}, \quad (9.15)$$

where y_i is a one-dimensional representation of the data sample \mathbf{x}_i and \mathbf{S}_{ij} is an entry in the similarity matrix \mathbf{S} that represent the adjacency graph. The adjacency graph weight α if nodes i and j are connected can be chosen by:

- A parameter function, $[t \in \mathbb{R}]$, e.g.

$$\alpha = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}}. \quad (9.16)$$

- A constant, e.g.

$$\alpha = 1. \quad (9.17)$$

Two ways of constructing the adjacency graph is:

- ϵ -neighborhood, $[\epsilon \in \mathbb{R}]$:

$$S_{ij} = \begin{cases} \alpha, & \|\mathbf{x}_i - \mathbf{x}_j\|^2 < \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (9.18)$$

- k nearest neighbors, $[k \in \mathbb{N}]$:

$$S_{ij} = \begin{cases} \alpha, & \text{if } \mathbf{x}_i \text{ is among the } k \text{ nearest neighbors of } \mathbf{x}_j \text{ or} \\ & \mathbf{x}_j \text{ is among the } k \text{ nearest neighbors of } \mathbf{x}_i \\ 0 & \text{otherwise} \end{cases} \quad (9.19)$$

The similarity matrix will inflict heavy penalties on the objective function in Eq. 9.15 if neighboring points \mathbf{x}_i and \mathbf{x}_j are mapped far apart in the output space. By minimizing the objective function LPP tries to ensure that y_i and y_j are close in the output space if \mathbf{x}_i and \mathbf{x}_j are close in the input space. The objective function can be reduced to

$$\arg \min_{\mathbf{w}} = \mathbf{w}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w} \text{ subject to } \mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w} = 1, \quad (9.20)$$

where the constraint $\mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w} = 1$ is used to normalize \mathbf{w} , by

$$\begin{aligned} & \frac{1}{2} \sum_{ij} (y_i - y_j)^2 s_{ij} && \Updownarrow \\ & \frac{1}{2} \sum_{ij} (\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j)^2 s_{ij} && \Updownarrow \\ & \sum_{ij} \mathbf{w}^T \mathbf{x}_i s_{ij} \mathbf{x}_i^T \mathbf{w} - \sum_{ij} \mathbf{w}^T \mathbf{x}_i s_{ij} \mathbf{x}_j^T \mathbf{w} && \Updownarrow \\ & \sum_i \mathbf{w}^T \mathbf{x}_i d_{ii} \mathbf{x}_i^T \mathbf{w} - \mathbf{w}^T \mathbf{X} \mathbf{S} \mathbf{X}^T \mathbf{w} && \Updownarrow, \\ & \mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w} - \mathbf{w}^T \mathbf{X} \mathbf{S} \mathbf{X}^T \mathbf{w} && \Updownarrow \\ & \mathbf{w}^T \mathbf{X} (\mathbf{D} - \mathbf{S}) \mathbf{X}^T \mathbf{w} && \Updownarrow \\ & \mathbf{w}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w} && \Updownarrow \end{aligned} \quad (9.21)$$

where \mathbf{X} is a matrix containing all data points $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, \mathbf{D} is a diagonal matrix where the elements is $\mathbf{D}_{ii} = \sum_j s_{ji}$ and $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is the Laplacian matrix. To normalize \mathbf{w} the system is constrained by $\mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w} = 1$.

The vector \mathbf{w} , which minimizes the objective function, is given by the eigenvector corresponding to the minimum eigenvalue solution of the generalized eigenvalue problem

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w} = \lambda \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w}. \quad (9.22)$$

If $\mathbf{X} \mathbf{D} \mathbf{X}^T$ has full rank it reduces the eigenvalue problem to

$$(\mathbf{X} \mathbf{D} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w} = \lambda \mathbf{w}. \quad (9.23)$$

Let $\mathbf{W} = [\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{n-1}]$ be the solutions to Eq. 9.23 ordered according to the eigenvalues, $\lambda_0 < \lambda_1 < \dots < \lambda_{n-1}$. Then the LPP projection can then be

written as

$$\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i. \quad (9.24)$$

It is worth noticing that the matrix $\mathbf{X}\mathbf{L}\mathbf{X}^T$ from LPP reduces to a PCA problem if ϵ from Eq. 9.18 is set to infinity. Thus the eigenvectors with the largest eigenvalues will correspond to the direction of maximal variance. Consult He *et al.* [30] for more insight on this.

9.3.1 LPP in Face Recognition Problems

Like FLDA, LPP has the problem of having n images consisting of d pixels where $n \ll d$. The matrix $\mathbf{X}\mathbf{D}\mathbf{X}^T$ is singular due to the fact that the rank of $\mathbf{X}\mathbf{D}\mathbf{X}^T$, which is an $d \times d$ matrix, is at most n .

As in FLDA this is usually solved by applying PCA² on the images prior to LPP, remove redundancy and make data more compact. The $\mathbf{X}\mathbf{D}\mathbf{X}^T$ matrix is made non-singular by only considering the t most important principal components from the PCA, where t is the number of non-zero eigenvalues of the $\mathbf{X}\mathbf{D}\mathbf{X}^T$ matrix.

9.4 Kernel Fisher Discriminant Analysis

Kernel Fisher Discriminant Analysis (KFDA) [44, 38] projects data in the input space into a high-dimensional feature space where a Fisher Linear Discriminant Analysis is performed on the data. Thus yielding a non-linear discriminant in the input space, as a result KFDA is like FLDA supervised. Let

$$\Psi : \mathbf{x} \in \mathbb{R}^n \rightarrow f \in \mathcal{F} \quad (9.25)$$

be the non-linear mapping from an d -dimensional input space to the e -dimensional implicit feature space \mathcal{F} , where $d \ll e$. The objective function of KFDA is sim-

²Normally capturing between 95-99% of the variance.

ilar to Eq. 9.12 of FLDA

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_B^\Psi \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W^\Psi \mathbf{w}}, \quad (9.26)$$

where $\mathbf{w} \in \mathcal{F}$ and the between- and within-matrices of \mathcal{F} are

$$\mathbf{S}_B^\Psi = \frac{1}{c(c-1)} \sum_{i=1}^c \sum_{j=1}^c (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j)(\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j)^T \quad (9.27)$$

and

$$\mathbf{S}_W^\Psi = \frac{1}{c} \sum_{i=1}^c \frac{1}{n_i} \sum_{j=1}^{n_i} (\Psi(\mathbf{x}_j) - \bar{\mathbf{u}}_i)(\Psi(\mathbf{x}_j) - \bar{\mathbf{u}}_i)^T, \quad (9.28)$$

where c is the number of classes, n_i is the number of samples in class i and $\bar{\mathbf{u}}_i$ denotes the mean of the samples in class i in \mathcal{F} given by

$$\bar{\mathbf{u}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \Psi(\mathbf{x}_j). \quad (9.29)$$

When the dimensionality of \mathcal{F} becomes very high (or even infinite), this problem becomes impossible to solve. However, by using the kernel trick [2] it is not necessary to compute Ψ explicitly. The kernel trick uses Mercer kernels to compute the inner product of two vectors in a feature space, \mathcal{L} , by using only the vectors of the input space, $K(\mathbf{x}, \mathbf{y}) = \Psi(\mathbf{x}) \cdot \Psi(\mathbf{y})$. A Mercer kernel satisfies the Mercer conditions:

- K is continuous.
- K is symmetrical, $K(\mathbf{x}, \mathbf{y}) = K(\mathbf{y}, \mathbf{x})$.
- K is positive semi-definite. That is for any set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ belonging to the input space, the matrix $\mathbf{M} = (K(\mathbf{x}_i, \mathbf{x}_j))_{ij}$, $\{i, j\} = 1, \dots, m$ is positive semi-definite.

The kernel trick states that any kernel $K(\mathbf{x}, \mathbf{y})$ satisfying the Mercer conditions can be expressed as a dot product in a high-dimensional space. Two examples in the Mercer kernels family are Gaussian kernels,

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}}, \quad t \in \mathbb{R}_+ \quad (9.30)$$

and polynomial kernels,

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x} \cdot \mathbf{y})^d, \quad d \in \mathbb{R}_+. \quad (9.31)$$

In order to avoid computing Ψ explicitly, Eq. 9.26 to Eq. 9.28 have to be formulated using only inner products of Ψ , i.e. $\Psi(\mathbf{x}) \cdot \Psi(\mathbf{y})$, which can be replaced by a kernel function.

Any solution $\mathbf{w} \in \mathcal{F}$ must lie in the span of all training samples in \mathcal{F} , proof of this can be obtained by consulting a reproducing kernels textbook. Therefore there exist n coefficients α_i so that an expansion of \mathbf{w} can be written as

$$\mathbf{w} = \sum_{i=1}^n \alpha_i \Psi(\mathbf{x}_i). \quad (9.32)$$

By use of Eq. 9.29 and Eq. 9.32 the projection of each class mean $\bar{\mathbf{u}}_i$ into \mathbf{w} can be written as

$$\mathbf{w}^T \bar{\mathbf{u}}_i = \boldsymbol{\alpha}^T \begin{bmatrix} \frac{1}{n_i} \sum_{j=1}^{n_i} K(\mathbf{x}_1, \mathbf{x}_j) \\ \frac{1}{n_i} \sum_{j=1}^{n_i} K(\mathbf{x}_2, \mathbf{x}_j) \\ \vdots \\ \frac{1}{n_i} \sum_{j=1}^{n_i} K(\mathbf{x}_n, \mathbf{x}_j) \end{bmatrix} = \boldsymbol{\alpha}^T \mathbf{m}_i. \quad (9.33)$$

Now, by use of Eq. 9.27 and Eq. 9.33 the numerator of Eq. 9.26 can be formulated as

$$\boldsymbol{\alpha}^T \mathbf{K}_B \boldsymbol{\alpha}, \quad (9.34)$$

where

$$\mathbf{K}_B = \frac{1}{c(c-1)} \sum_{i=1}^c \sum_{j=1}^c (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T. \quad (9.35)$$

Similarly, by use of Eq. 9.28, Eq. 9.29 and Eq. 9.32 the denominator of Eq. 9.26 can be formulated as

$$\boldsymbol{\alpha}^T \mathbf{K}_W \boldsymbol{\alpha}, \quad (9.36)$$

where

$$\mathbf{K}_W = \frac{1}{c} \sum_{i=1}^c \frac{1}{n_i} \sum_{j=1}^{n_i} (\boldsymbol{\zeta}_j - \mathbf{m}_i)(\boldsymbol{\zeta}_j - \mathbf{m}_i)^T \quad (9.37)$$

and

$$\boldsymbol{\zeta}_j = [K(\mathbf{x}_1, \mathbf{x}_j) \ K(\mathbf{x}_2, \mathbf{x}_j) \ \cdots \ K(\mathbf{x}_n, \mathbf{x}_j)]^T. \quad (9.38)$$

Eq. 9.26 can be rewritten using Eq. 9.34 and Eq. 9.36 yielding

$$\max_{\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}^T \mathbf{K}_B \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \mathbf{K}_W \boldsymbol{\alpha}}. \quad (9.39)$$

FLDA can now be performed implicit in \mathcal{F} by calculating the leading eigenvectors of $\mathbf{K}_W^{-1} \mathbf{K}_B$ in the input space. The projection of a given point \mathbf{x} into \mathbf{w} in \mathcal{F} is given by

$$\mathbf{w} \cdot \Psi(\mathbf{x}) = \sum_{i=1}^n \alpha_i K(\mathbf{x}_i, \mathbf{x}). \quad (9.40)$$

9.4.1 Problems of KFDA

KFDA has numerical problems since \mathbf{K}_W is seldom of full rank. A simple way to avoid this is to add a multiple of the identity matrix to \mathbf{K}_W [44] so that

$$\tilde{\mathbf{K}}_W = \mathbf{K}_W + \mu I. \quad (9.41)$$

Another problem is how to select a kernel function for different tasks, which is still an open problem [38].

CHAPTER 10

Experimental Results I

The purpose of this chapter is to evaluate the the four feature extraction methods described in Chapter 9. To summarize, the four extraction methods are:

- Principal Component Analysis
- Locality Preserving Projections
- Fisher Linear Discriminant Analysis
- Kernel Fisher Discriminant Analysis

The idea is to extract facial features from facial images by use of the four methods mentioned above presuming that the facial features are easily separated in the feature space. For the tests presented in this chapter `data set I` is used. The tests are divided into two parts, where the first is an illustration of the four feature spaces, while the second tests face identification using a nearest neighbor classifier in the four output spaces. The choice of nearest neighbor algorithm as classifier used in the second part is based on the following observations:

- It is very simple, yet powerful if it contains an adequate number of classified samples, resulting in well defined populations.

- It does not need training.
- It produces reproducible results, i.e. if several tests are conducted using the same protocol for testing the obtained results are unambiguous. In contrast, neural networks use some degree of randomness in the training, which in some cases results in non-reproducible results.

Eq. 9.17 ($\alpha = 1$) is used in these tests as the entries in the LPP's similarity matrix and it is constructed according to Eq. 9.19 (the k nearest neighbors). After careful consideration $k = \lfloor m/2 \rfloor$ was used, since this seems to yield the best results. Here, m is the average number of samples in the training set of a specific person. For KFDA the Gaussian kernel in Eq. 9.30 is used, where $t = 250000$ was used, since this seemed to yield the best results.

10.1 Illustration of the Feature Spaces

In the following sections feature spaces of PCA, LPP, FLDA and KFDA are illustrated by scatter plot matrices of the three most discriminative dimensions. **Data set I** consisting of 120 images of 12 persons was divided into two groups of 60 images each. The first group consists of the images $\{1, 3, 5, 7, 9\}$ of each person¹ and is used to calculate the mapping between the image space and the feature space. The second group consists of the images $\{2, 4, 6, 8, 10\}$. These are mapped into the feature spaces to visualize the separability of the 12 persons when using the four methods. Furthermore, the test was conducted to obtain a combined projection of photometric and geometric information. Tests could be conducted only concerning either the photometric or the geometric information, but due to space constraints these tests were not included.

The PCA space is constructed like a AAM (described in Section 7.2.5). First a PCA is performed on the shape free images and then on the shapes. The obtained results were then combined and a third PCA was performed yielding a subspace containing both the geometric and photometric information. When combining the results obtained from the first two PCA's the weighting between texture and shape is one (i.e. $\mathbf{W}_s = \mathbf{I}$ in the AAM terminology) and the variance captured by the PCA accounts for 99.9%. The PCA scatter plot matrix obtained in the test is displayed in Figure 10.1.

The LPP, FLDA and KFDA spaces are constructed in the compact PCA space described above. This is done to avoid singularity problems (described in Chapter 9) and to reduce the huge workload associated with the image space of

¹The image numbers are described in the technical rapport in Appendix A.

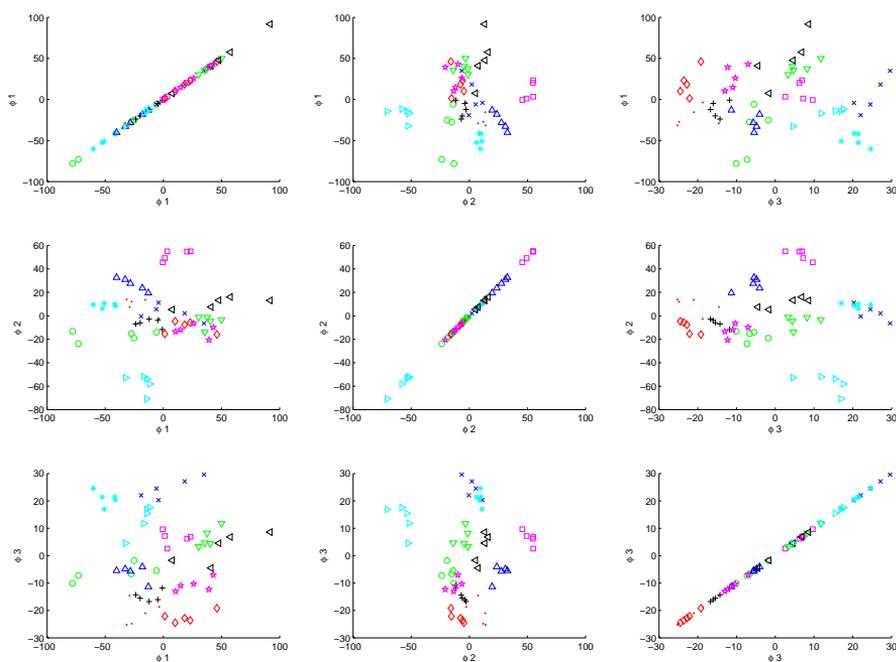


Figure 10.1: The PCA scatter plot matrix of the combined features from data set I. The 12 persons of data set I are displayed by different colors and symbols.

KFDA. The LPP, FLDA and KFDA scatter plot matrices are displayed in Figure 10.2, Figure 10.3 and Figure 10.4, respectively.

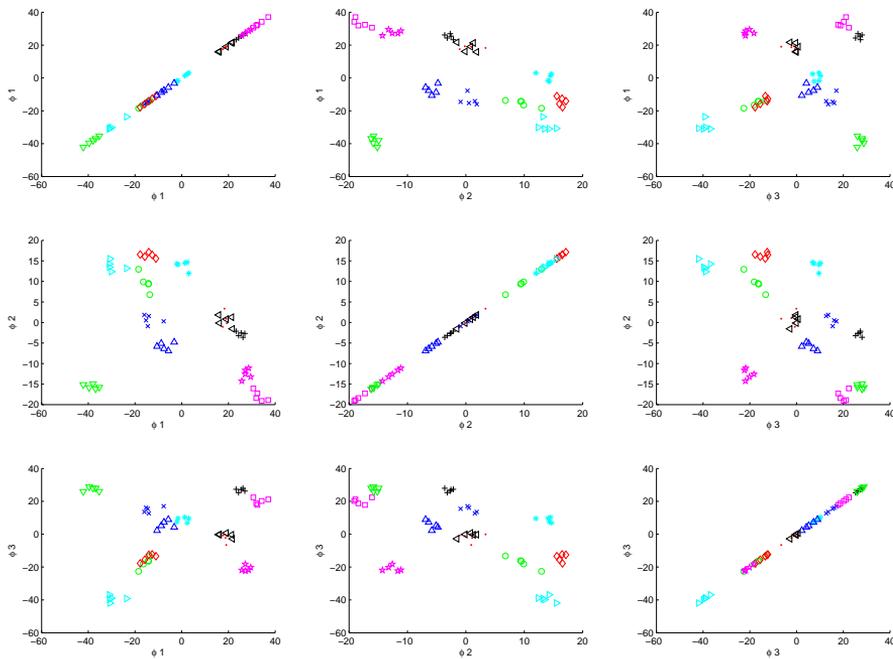


Figure 10.2: The LPP scatter plot matrix of the combined features from data set I. The 12 persons of data set I are displayed by different colors and symbols.

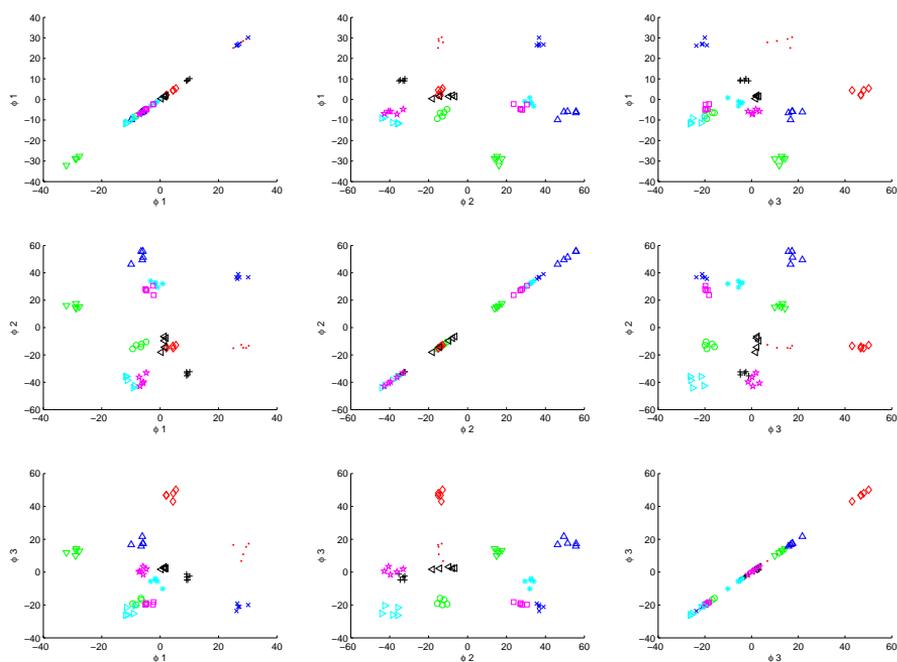


Figure 10.3: The FLDA scatter plot matrix of the combined features from data set I. The 12 persons of data set I are displayed by different colors and symbols.

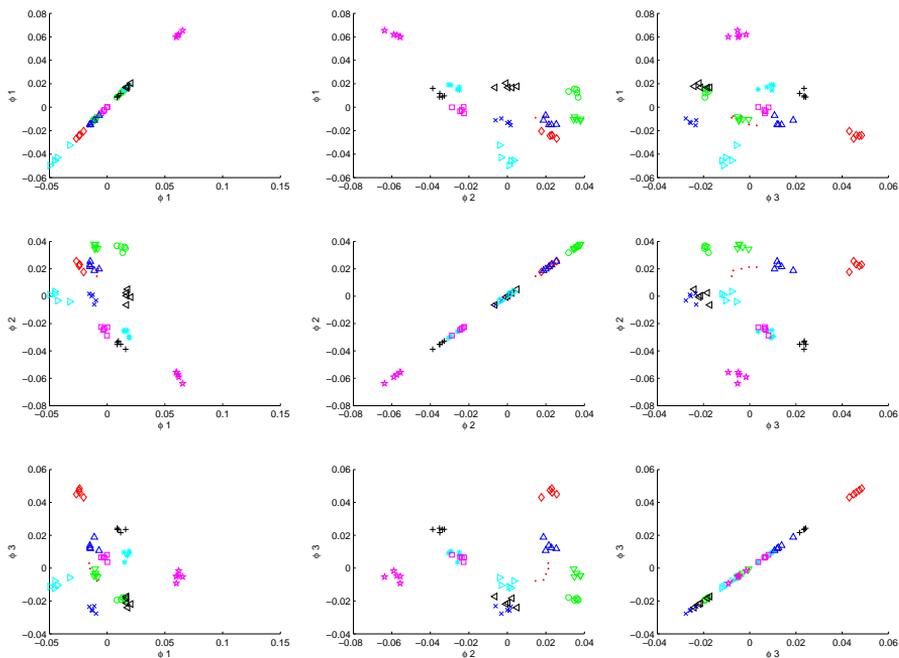


Figure 10.4: The KFDA scatter plot matrix of the combined features from data set I. The 12 persons of data set I are displayed by different colors and symbols.

10.2 Face Identification Tests

The following face identification tests are conducted by dividing `data set I` into a training set and a test set. The training set is used to construct the feature space (as described in the previous section) as well as used by the nearest neighbor algorithm to classify new samples. The test set is classified by the nearest neighbor algorithm from which the false identification rates are calculated.

10.2.1 50/50 Test

The 50/50 test uses the same scheme of dividing `data set I` into training set and test set described in Section 10.1. The false identification rates are display in Figure 10.5.

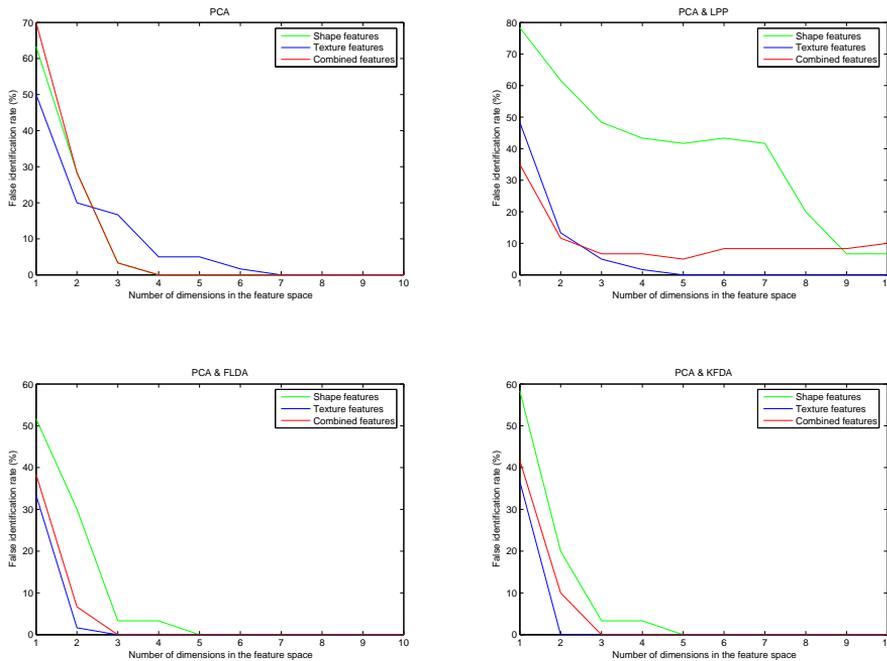


Figure 10.5: False identification rates obtained from the 50/50 test, where `data set I` is divided into two equal sized sets, a training set and a test set. The false identification rates are plotted as a function of the feature space dimension.

10.2.2 Ten-fold Cross-validation Test

In the ten-fold cross-validation test **data set I** is divided into ten folds, each fold consisting of one image of all persons. The cross-validation test was performed at each of the ten iterations by using nine folds as the training set while using the remaining fold as test set. The false identification rates are display in Figure 10.6.

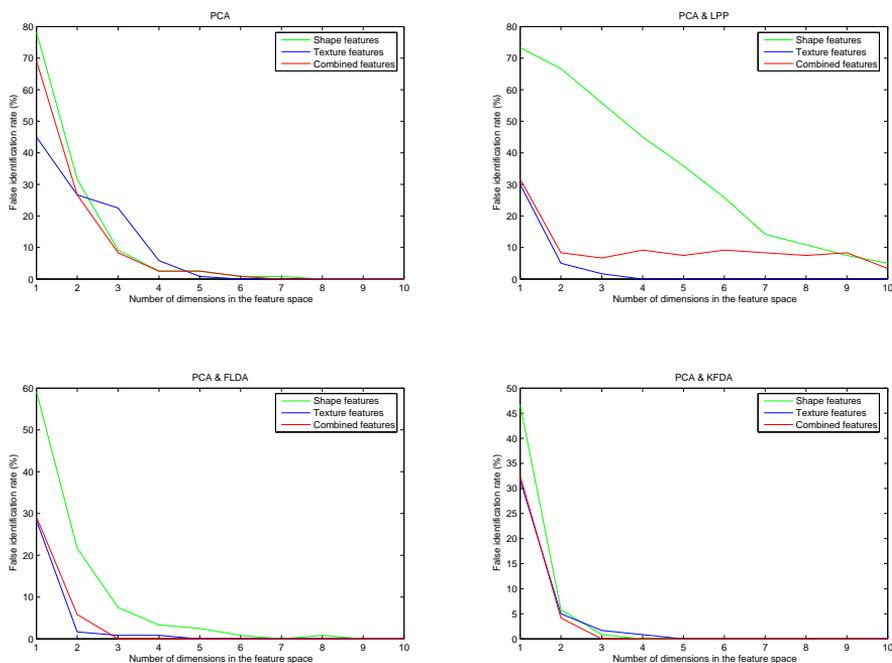


Figure 10.6: The false identification rates obtained from the ten-fold cross-validation test. The false identification rates are plotted as a function of the feature space dimension.

10.3 Discussion

Analysis of the scatter plot matrices, displayed in Figure 10.1, Figure 10.2, Figure 10.3 and Figure 10.4 indicate that the methods LPP, FLDA and KFDA improve the separability of the classes (persons) compared to PCA. Though, LPP seems to have a problem separating two of the classes. This is supported by the LPP FIR graphs (in Figure 10.5 and Figure 10.6) where the combined feature shows a general FIR of 10%. In addition the shape feature display a very high percentage of FIR in an interval of 10 to 80%.

These problems with the shape and combined features displayed by LPP are probably due to the way the similarity matrix is constructed. In the current implementation the nearness in the input space is defined by Euclidean distances, which is not necessarily a meaningful measurement of a shape.

When using KFDA one has to consider the choice of a kernel function resulting in a satisfying result. This choice is as described earlier still an open problem. The parameter t in the chosen kernel function used in the presented tests has to be adjusted when the KFDA method is used on a new data set. Another problem when using KFDA compared to linear methods is the higher risk of overfitting the training data, which in turn causes a drop in performance. KFDA as a result can be considered “too flexible” from an implementation point of view.

Analysis of the FIR graphs shows that FLDA performs nearly as well as KFDA. FLDA, in contrast to LPP and KFDA, does not contain any parameters needing initializing, which makes FLDA a very robust method. Therefore, the design of a new face recognition algorithm presented in the next chapter is built on FLDA.

The problem of improving the similarity matrix in LPP is not pursued further in this thesis.

The overall results show that the variation of data set I is very small.

Part III

Development

Multiple Individual Discriminative Models

Typically, the techniques PCA, LPP, FLDA and KFDA are in the literature classified with a nearest neighbor algorithm [40, 29, 44].

However, some inconvenience appears when testing the techniques using unknown persons in the test set. In this case, a criterium has to be chosen to decide whether or not an image of an specific person belongs to the training set, e.g. only people with a Euclidean distance less than a given threshold are considered as part of the training set. However, such a threshold should not necessarily be the same for all classes (persons). The estimation of different thresholds is not straightforward. Additional data are needed to estimate these. In addition, it is a problem to determine the optimal dimensionality of the four techniques described above.

In this chapter, a novel technique is proposed that addresses the inconveniences described above. In this technique, denoted Multiple Individual Discriminative Models (MIDM), the data are projected by FLDA into c one-dimensional spaces instead of the “normal” FLDA $(c - 1)$ -dimensional space. Here, c is the number of people in the data set.

The aim of each of the individual models, obtained by the projection of the training set into the c one-dimensional spaces, is to characterize a given person

uniquely, resulting in each of the individual models representing one person. The multiple one-dimensional models allow statistical interpretation of the “degree of membership” of a given person and allows detection of unknown persons (faces). Furthermore, these individual models have several advantages in interpretability, characterization, accuracy and easiness in updating one or more models belonging to the MIDM.

11.1 Algorithm Description

The proposed algorithm is composed of two steps. The first step builds an individual model for each person in the database using the photometrical and geometrical information provided by the available images. Each model characterizes a given person and discriminates the person from the other people in the database. The second step carries out the identification. A classifier, which is related to the standard Gaussian distribution, decides whether or not a facial image belongs to a person in the database. In this section, the two steps of the algorithm are described in detail. A diagram of the algorithm is displayed in Figure 11.1. This diagram will be referred to during the description of the algorithm to allow the reader to gain a deeper understanding.

11.1.1 Creations of the Individual Models

11.1.1.1 Obtaining the Geometry of the Face

The geometrical characterization of a given face is obtained by means of the theory of statistical shape analysis [18]. In this theory, objects (faces) are represented by shapes.

Definition 11.1 *A shape is, according to Kendall [33], all the geometrical information that remains when location, scale and rotational effects are filtered out from an object.*

In order to describe a shape, a set of landmarks or points of correspondence matching between and within populations are placed on each face.

To obtain a shape representation according to the Definition 11.1, the obtained landmarks are aligned in order to remove the location, rotational and scaling effects. To achieve this goal, the 2D Procrustes analysis is carried out (Figure 11.1 A).

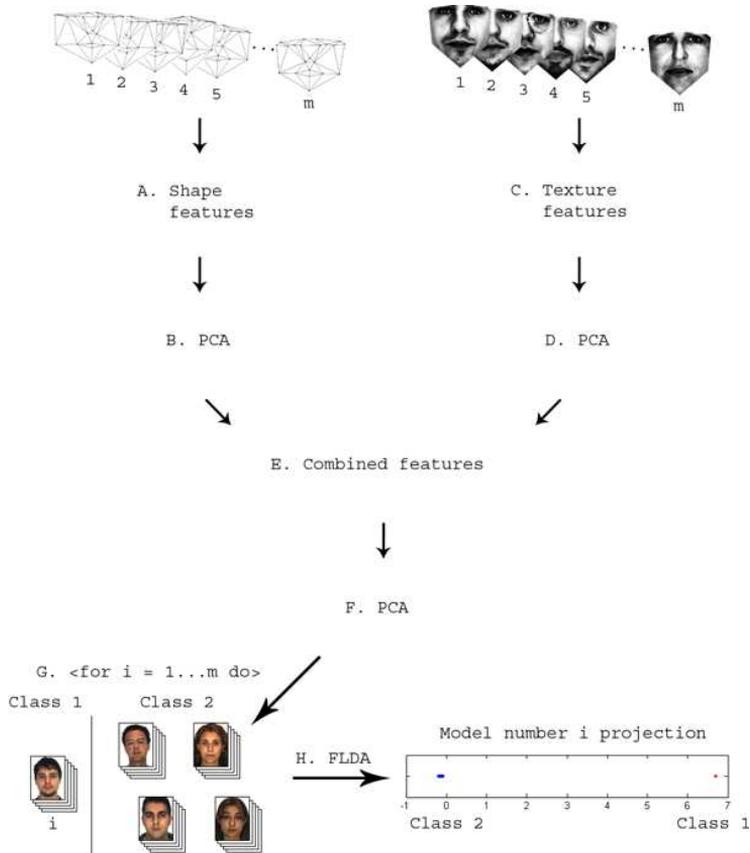


Figure 11.1: Overview of the MIDM algorithm. A: Landmark alignment using full Procrustes analysis. B: PCA conducted on aligned landmarks to remove redundancy. C: Texture normalization using global histogram equalization. D: PCA conducted on normalized texture to remove redundancy. E: Combination of shape and texture feature information. F: PCA conducted on combined features to remove redundancy. G & H: Process of building the individual model using FLDA, where each person in turn is left out of class 2 making the projection that discriminates that person.

In order to remove redundancy in the shape data, PCA is applied to the aligned landmarks (Figure 11.1 B).

11.1.1.2 Texture Formulation

To form a complete model of the face appearance the MIDM algorithm captures the photometric information provided by pixels. In order to collect this texture representation a Delaunay triangulation [46] of every shape is obtained. The Delaunay triangulation connects the aligned landmark set of each image by a mesh of triangles, so no triangle has any of the other points of the landmark set inside its circumcircle. The Delaunay triangulation of the landmarks scheme in **data set II** is displayed in Figure 11.2.

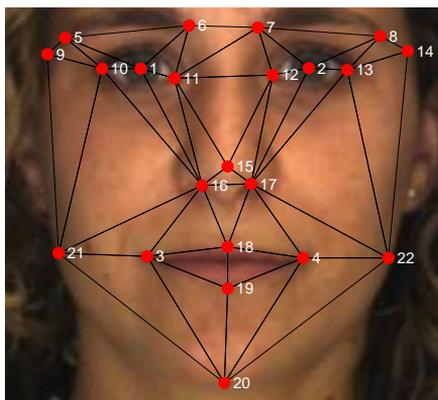


Figure 11.2: Example of a Delaunay triangulation of the landmark scheme in **data set II**.

The Delaunay triangulation obtained for each image is warped onto the Delaunay triangulation of the mean shape. Formally, let I be a given image and M the mean shape previously obtained. Let $\mathbf{u}_1 = [x_1, y_1]$, $\mathbf{u}_2 = [x_2, y_2]$ and $\mathbf{u}_3 = [x_3, y_3]$ denote the vertices of a triangle T in I , and let $\mathbf{v}_1, \mathbf{v}_2$ and \mathbf{v}_3 be the associated vertices of the corresponding triangle in M . Given any internal point $\hat{\mathbf{u}} = [x, y]$ in the triangle T , the corresponding point in the associated triangle in the mean shape can be written as $\hat{\mathbf{v}} = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2 + \gamma\mathbf{v}_3$ where

$$\alpha = 1 - (\beta + \gamma), \quad (11.1)$$

$$\beta = \frac{yx_3 - x_1y - x_3y_1 - y_3x + x_1y_3 + xy_1}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2}, \quad (11.2)$$

$$\gamma = \frac{xy_2 - xy_1 - x_1y_2 - x_2y + x_2y_1 + x_1y}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2}. \quad (11.3)$$

This transformation extracts the same shape free image of a given face image as produced by AAM in Section 7.2.5. A histogram equalization is applied to the collected texture as described in Section 8.1.1 to reduce the effects of illumination differences. This histogram equalization is performed independently in each of the three color channels (RGB) to reduce lightning artifacts. Afterwards, the three color channels can be converted into gray scale if a more compact representation is desirable (Figure 11.1 C). PCA is conducted on the texture data to reduce dimensionality and data redundancy (Figure 11.1 D). The solution to the memory problems due to the huge dimension of the covariance matrix is solved as described in Section 9.1.2.

11.1.1.3 Combining Color and Geometry

The shape and texture features are combined in a matrix (Figure 11.1 E). In order to remove correlation between shape and texture and to make the data representation more compact a third PCA is performed on the combined shape and texture matrix (Figure 11.1 F).

11.1.1.4 Building an Individual Model

Once the geometrical and photometrical information of the faces has been captured, the MIDM algorithm builds an individual model for each person in the training set. Each individual model is built in two steps. First the training set is divided into two groups, one representing a person i and one representing persons other than i (Figure 11.1 G). Second FLDA is applied on the two groups (Figure 11.1 H). This is repeated for every person in the training set, resulting in individual models representing each person.

11.1.2 Classification

In order to obtain a method to classify a given image, the different individual models are first standardized. The standardization of model $i = 1, \dots, m$ is based on two assumptions. First, the number of observations for person i is much smaller than the number of the observations for all other people. Second the projection of the other people follows a Gaussian distribution. The two assumptions imply that the distribution of all the observations can be assumed as a Gaussian distribution with outliers. The standardization of model i is then achieved by transforming the projections into a standard Gaussian distribution, keeping the projections of the person i positive. Formally, let \bar{x}_i be the mean of the projections on model i , σ_i the standard deviation and $x_{i,j}$ the projection of image j in model i . These projections are standardized by

$$\hat{x}_{i,j} = (x_{i,j} - \bar{x}_i) / \sigma_i. \quad (11.4)$$

If the standardized projections of the images corresponding to person i are negative, then $\hat{x}_{i,j}$ are replaced by $-\hat{x}_{i,j}$ for all projections. This causes the projections of the images corresponding to person i to be positive and far from the mean of the Gaussian distribution.

Once the model i is standardized, the probability that a projected image belongs to the person i is given by the value of the standard normal cumulative function in the projected value. This is used to classify a given image. If it is assumed that the image belongs to a person from the training set, the image is projected by all the individual models of MIDM. The image is then classified as belonging to the model that results in the largest probability. Furthermore, it is statistically possible to decide whether or not a given person belongs to the training set. This can be achieved by comparing the largest projection obtained in all the models with a threshold. E.g. if a probability of 99.9% is required, a given image will only be considered as belonging to the database, if the projection in one of the individual models is higher than 3.1 standard deviations. This in turn estimates the FAR to be 0.1%. It is not possible to estimate the FRR this way. The estimation of FRR, like other classifiers, requires use of test data set.

11.2 Discussion

The MIDM technique is very robust and highly accurate (as shown later in this thesis) and it is based on a very simple setup analyzing single 2D images. Furthermore, the projected subspaces yield a very intuitive interpretation, i.e.

a high projected value indicates a large probability of that image belonging to the person that the model represents. The statistical classifier does not need training. Instead the value of the threshold can be selected to achieve a desired FAR, i.e. the FAR is $1 - \mathcal{P}(t)$ where \mathcal{P} is the cumulative distribution function of the standard normal distribution and t is the threshold in standard deviations. However, it is possible to train the classifier if e.g. the optimal separation of the imposter and the client classes is of interest (EER threshold). The performance of MIDM is extensively reported in Chapter 13.

Reconstruction of 3D Face from a 2D Image

In order to achieve better results in the face recognition process, face recognition in three dimensions would be interesting, since the face is a three-dimensional object. 3D data can be collected in various ways, as already described in Chapter 4. However, using the knowledge of statistical shape analysis it could be very interesting to investigate the extraction of 3D information from a 2D image using prior knowledge of faces in 3D.

This chapter will consult this issue, using only frontal facial images. Since the topic is on the edge of the scope of this thesis it will not be a conclusive presentation of the area, but only a brief appetizer.

12.1 Algorithm Description

Extraction of the 3D information from a 2D image is based on the existence of a statistical shape model of 3D faces, i.e. a Point Distribution Model (PDM) [11]. By use of PDM the information obtained from a 2D image of a face can be used to reconstruct the third image dimension, by manipulating the PDM until an optimal fit is achieved. The fit is calculated by projecting the already annotated

landmarks in the 3D shape into 2D followed by calculation of the least-squares distance to the 2D image landmarks. An illustration of this process is displayed in Figure 12.1.

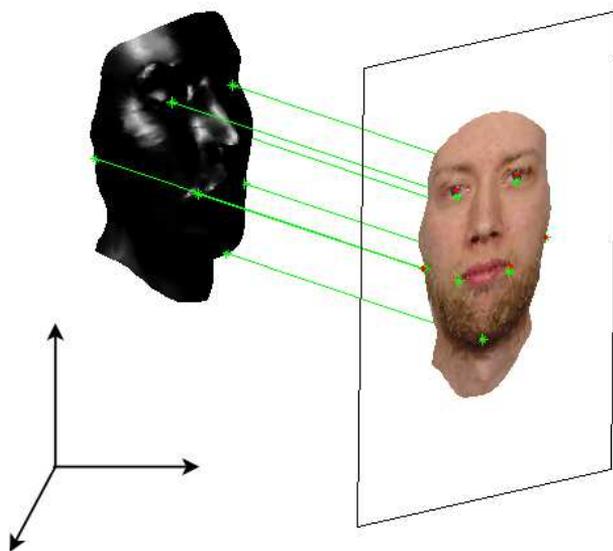
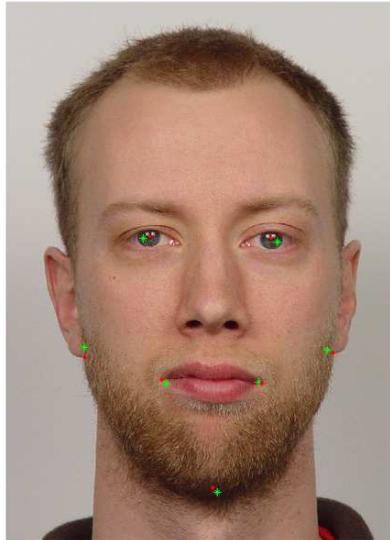


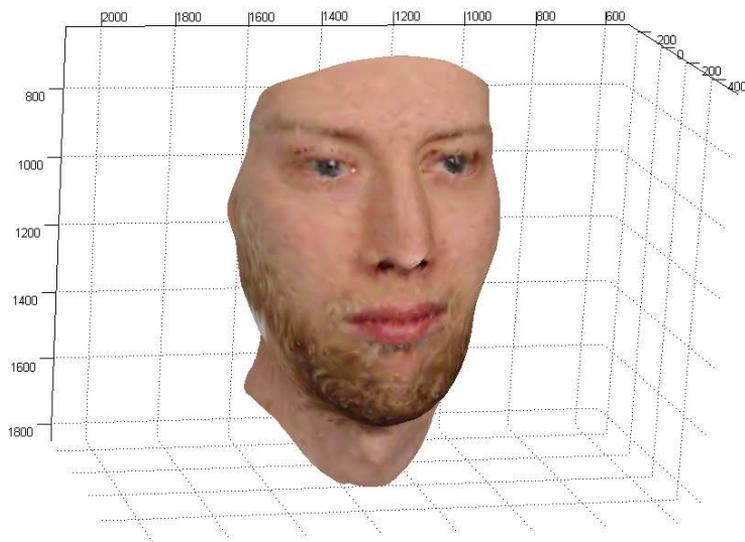
Figure 12.1: Process of projecting 3D landmarks into a 2D image, which is used to calculate a fit of the 3D shape.

Since the PDM algorithm is only used on frontal images, the problem of estimating the pose of the 2D image to obtain the correct projection of the 3D model is not considered here. The PDM is constructed from **data set IV** described in Chapter 6 and is annotated by a seven landmark scheme (a subset of the landmark scheme used in **data set I**). An example of depth retrieval from one 2D image is displayed in Figure 12.2.

The PDM is fitted to landmarks in 2D. However, features hidden by the pose of the 2D image may not be reconstructed correctly in 3D. An example is the curvature of the nose. It can not be seen from a 2D frontal image whether the nose is convex or concave. This can be solved by analyzing multiple 2D images of different facial poses. An example of this is displayed in Figure 12.3.



(a)



(b)

Figure 12.2: Example of retrieved depth information from a 2D frontal face image. (a) The 2D and 3D landmarks superimposed on the frontal face. The 2D and 3D landmarks are plotted in red and green, respectively. (b) The obtained 3D representation of the 2D image shown in (a).

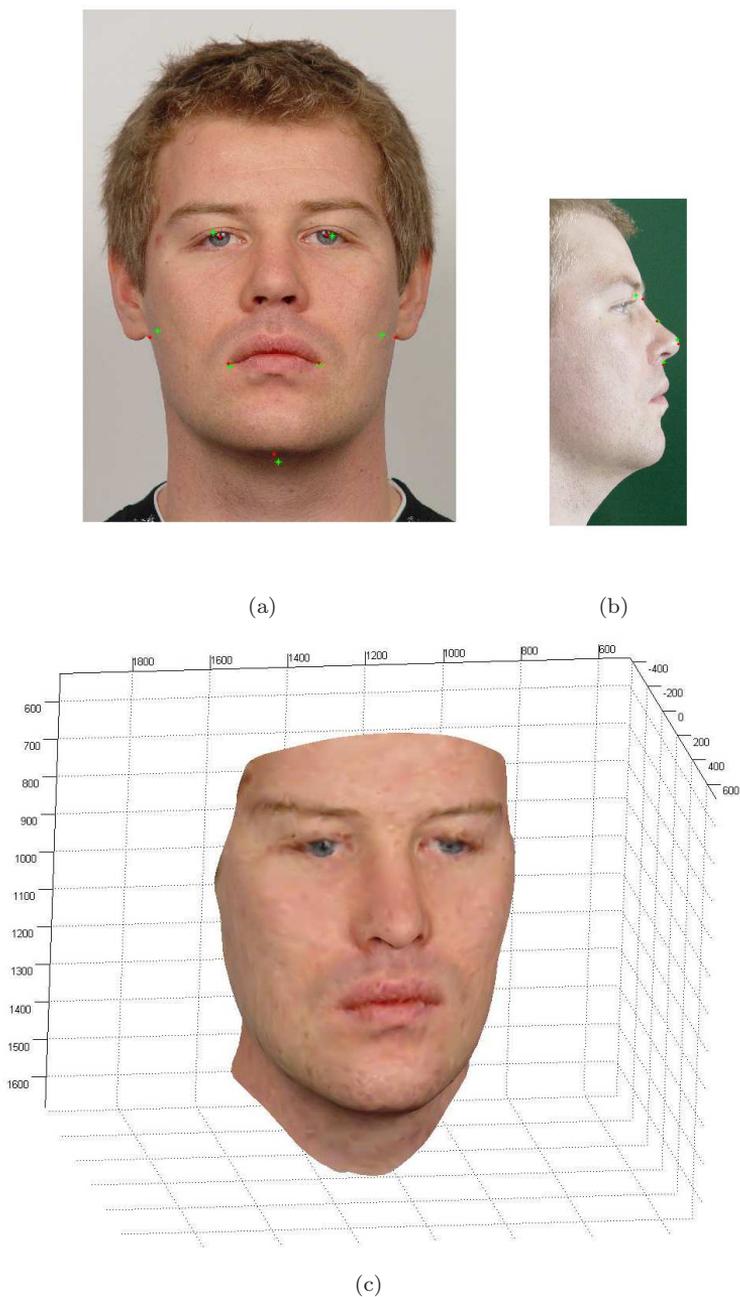


Figure 12.3: Example of retrieved depth information from a 2D frontal face image and a 2D profile face image. (a) and (b) The 2D and 3D landmarks superimposed on the frontal face and a profile face image, respectively. The 2D and 3D landmarks are plotted in red and green, respectively. (c) The obtained 3D representation of the two 2D images shown in (a) and (b).

12.2 Discussion

No testing was performed to analyze how much the recognition is improved by reconstructing a 3D image from a 2D image. **Data set IV** only contains 24 individuals, and as a result it does not contain enough variation to capture new faces differently from the images of **data set IV**. However, the reconstructions performed are very promising, and 3D reconstruction is definitely an area of interest when performing future work on face recognition.

CHAPTER 13

Experimental Results II

13.1 Overview

In this chapter an evaluation of the MIDM algorithm proposed in Chapter 11 is performed. The evaluation is divided into two parts as described in the following:

13.1.1 Initial Evaluation Tests

The initial evaluation will be done by performing the following tests on the MIDM algorithm using `data set II`:

- **Identification test.** This test will show the performance of MIDM against a regular Fisherface with a nearest neighbor classifier. In addition, this experiment aims at testing whether or not the shape information is important to the recognition process.
- **Verification test.** In this test a 25-fold cross-validation scheme is used to ascertain the FAR and FRR for `data set II` of the MIDM.

- **Robustness test.** By superimposing spectacles on a group of people it is tested whether or not these persons identities obtained by MIDM change.

Furthermore, the pixels most important in the recognition process are identified for selected models (persons) in the MIDM.

13.1.2 Lausanne Performance Tests

These following tests are conducted on the MIDM algorithm using **data set III** according to the Lausanne protocol [39], and the obtained results are to be used in the participation in the ICBA2006 Face Verification Contest in Hong Kong in January, 2006:

- A **Verification test** using the Lausanne protocol configuration I with manual annotation of landmarks.
- A **Verification test** using the Lausanne protocol configuration II with manual annotation of landmarks.
- A **Verification test** using the Lausanne protocol configuration I with automatic annotation of landmarks.
- A **Verification test** using the Lausanne protocol configuration II with automatic annotation of landmarks.

The obtained results will be compared with results reported in the AVBPA2003¹ Face Verification Contest, 2003.

13.2 Initial Evaluation Tests

Data set II is used for all the initial evaluation tests.

13.2.1 Identification Test

The first test conducted is an identification test, testing the performance of the MIDM method with respect to the Fisherfaces method in terms of correct

¹The Audio Video Biometric Person Authentication Conference, 2003.

classification rates.

In order to evaluate the importance of the geometrical information, the Fisher-face technique was modified by replacing the texture data with the shape data. In addition, the shape data and the texture data were combined. The two modified techniques are referred to as Fishershape and Fishercombined from now on. In addition, the three Fisher-methods are implemented using the shape free images instead of normal images, which are generally used by Fisherfaces. This will make the tests more compatible, as they use the same data.

The Euclidean Nearest Neighbor algorithm was used as classifier in the Fisher methods. The MIDM method classified images as belonging to the model yielding the highest probability of recognition as described in Section 11.1.2.

The images from **data set II** were divided into two sets. Images from the first session were used to build the individual models, and images from the second session were subsequently used to test the performance.

The MIDM method was built using the geometrical (shape), the photometrical (texture) and the combined information of each face corresponding to Fishershape, Fisherface and Fishercombined, respectively.

The test was repeated exchanging the roles of the training and test sets, i.e. images from session two was used as training set and images from session one as test set. The correct identification rates from using the different techniques are shown in Table 13.1.

The MIDM method shows a higher rate of performance than the Fisher methods (Table 13.1). Moreover, by analyzing the obtained results in Table 13.1 it is indicated that by use of texture data one obtains a higher accuracy than by using shape data. This implies that the information contained in the texture data is more significant than the information included in the shape data. However, the information contained in the shape data is not insignificant, since the highest correct classification rate is obtained when combining both shape and texture.

The importance of geometrical information in facial recognition can be further substantiated by retrieving the shape free image of a person from **data set I** who has a very slim face. When looking at the normal image and the shape free image displayed in Figure 13.1 even humans can have difficulties determining whether or not the two images are of the same person. Though, the difference between an shape and the mean shape is not always as significant as displayed

²Number of misclassified images reported in parentheses.

³The Fisher-methods are modified to use shape free images.

Method	Input features	Correct Identification Rate ²
MIDM method	Shape	86.4% (95)
MIDM method	Texture	99.6% (3)
MIDM method	Texture and Shape	99.9% (1)
Fishershape ³	Shape	85.9% (99)
Fisherface ³	Texture	96.9% (22)
Fishercombined ³	Texture and Shape	97.1% (20)

Table 13.1: Correct identification rates.

in Figure 13.1.

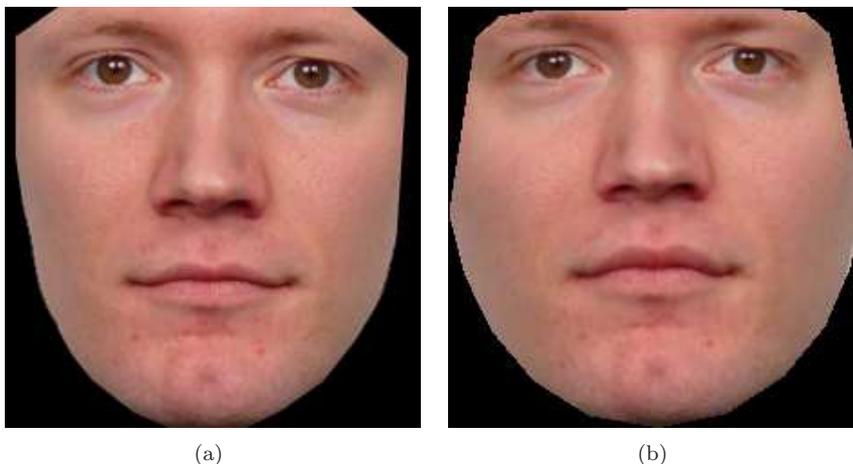


Figure 13.1: Importance of geometrical information. (a) the facial region of an image in data set I, (b) the facial region from (a) displayed as a shape free image (warped into the meanshape).

13.2.2 The Important Image Regions

An interesting property of the MIDM algorithm is that it is capable of determining the most discriminative features of a given person. To illustrate this, four models were built using only the texture information. The pixels of the faces corresponding to these models are displayed in Figure 13.2. Here the 10%, 15% and 25% pixels of the weights of highest value are displayed in red. By analyzing Figure 13.2 it is clear that the most important discriminating features include eyes, noses, spectacles, moles and beards. Notice that the MIDM algorithm detects the spectacles and the two moles of person 43 as discriminating features.

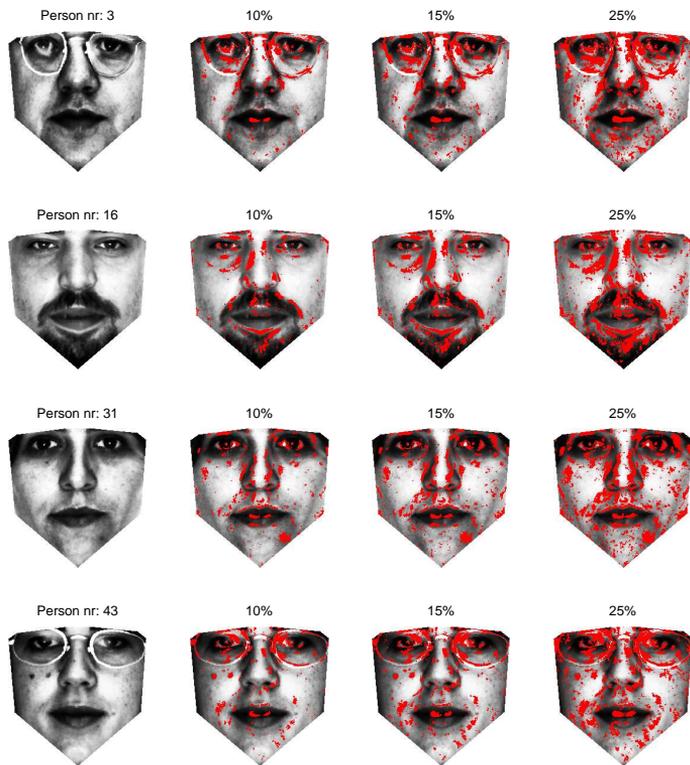


Figure 13.2: The 10%, 15% and 25% pixels weights of highest value most important for discrimination between the test persons of data set II. The mentioned pixels are shown in red.

13.2.3 Verification Test

A verification test was performed to verify whether or not the MIDM algorithm can be incorporated into a biometrical facial recognition scheme, i.e. whether or not MIDM can identify people of the training set as well as detecting unknown people.

A 25-fold cross-validation scheme was conducted in order to evaluate the performance of MIDM. Seven face images of one male and seven face images of one female of `data set II` were left out in each cross-validation iteration, i.e. images of 48 people were used in the training of the MIDM algorithm. Just to summarize, a total of seven images per person were included in the training set. The “left out” male and female were considered as imposters (unknowns).

The verification test was conducted in two scenarios:

- **Best case scenario.**

In the best case scenario the imposter is obligated to try out all known identities in the training set in the attempt to be recognized by the MIDM. This is necessary since the imposters in the best case scenario does not know which person in the training set he/she resembles most. The 25-fold cross-validation results in 8400 client validations ($25 \text{ folds} \times 48 \text{ persons} \times 7 \text{ images}$) and 16800 imposter validations ($25 \text{ folds} \times 2 \text{ persons} \times 48 \text{ total number of identities} \times 7 \text{ images}$).

- **Worst case scenario.**

In the worst case scenario the imposters do know which person in the training set they reassemble most. The imposters will therefore attempt to be recognized by the MIDM only using this identity. In this scenario the 25-fold cross-validation results in 8400 client validations ($25 \text{ folds} \times 48 \text{ persons} \times 7 \text{ images}$) and 350 imposter validations ($25 \text{ folds} \times 2 \text{ persons} \times 7 \text{ images}$).

The average False Acceptance Rate (FAR) and average False Rejection Rate (FRR) graphs obtained in the best and worst case scenarios as well as the corresponding average Receiver Operating Characteristic curve (ROC) are displayed in Figure 13.3.

The verification test shows a high degree of separability between the client and imposter populations for both scenarios. The highest degree of separation occurs at the Equal Error Rate, yielding FAR and FRR at 0.3% and 1.8% obtained for the best and worst case scenario, respectively.

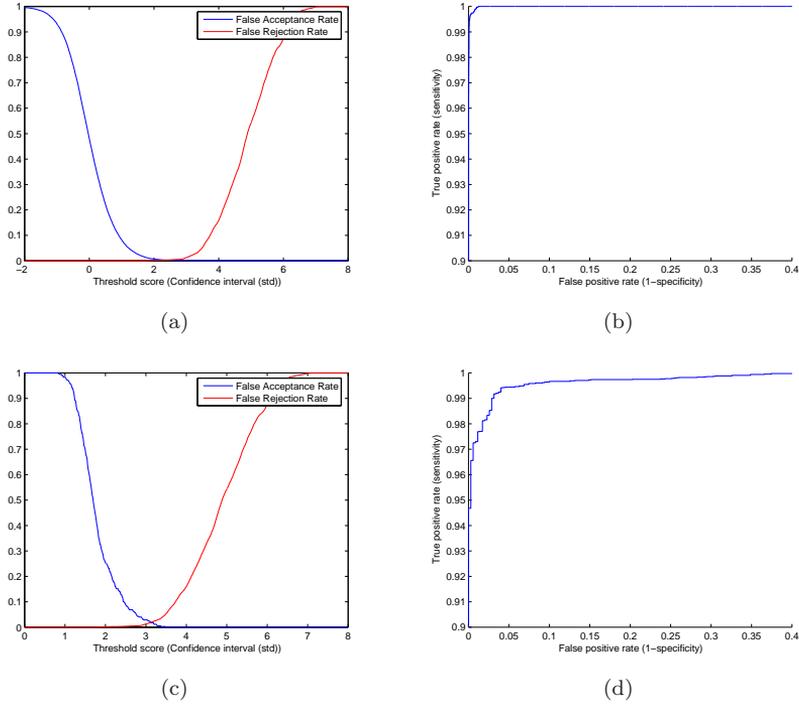


Figure 13.3: Best case (a. & b.) and worst case (c. & d.) scenario verification tests. (a) & (c) False Acceptance Rate and False Rejections Rate graphs obtained by the 25-fold cross-validation for the best and worst case scenario, respectively. (b) & (d) Receiver Operating Characteristic (ROC) curve obtained by the 25-fold cross-validation for the best and worst case scenario, respectively. Notice that only the top left part of the ROC curves is displayed.

In addition, if the algorithm belongs to a security scheme, the degree of accessibility can be changed by increasing or decreasing the threshold of the classifier, e.g. if a false rejection rate of 5.5% is allowed in the worst case scenario a 0% false acceptance rate is obtained. This accommodates biometrical security systems requiring a high level of access control.

13.2.4 Robustness Test

Finally, a test is conducted in order to estimate the robustness of the MIDM method. In addition, the test aims at settling that the MIDM method not only discriminates on removable features, such as spectacles, but uses permanent facial features in the discrimination process, as well. To achieve this goal eight people (four male and four female) are synthetically fitted with four different types of spectacles. The spectacles were obtained from people belonging to **data set II**. By fitting the eight people with the four types of spectacles 32 synthetic images was obtained.

The test was conducted in two steps.

- First, all images in session of **data set II**, except for images of the eight selected persons, were used to build the MIDM. The goal is to examine whether or not the eight selected people are considered as belonging to the training set. The results show that none of the 32 synthetic images are misclassified when considering a threshold of 3.1 standard deviations i.e. the FAR will approximately be 0.1%. This is displayed in Figure 13.4 II, where the projections of one of the eight unknown people into the different models of the MIDM are displayed. The figure shows, that when the displayed person is unknown to the MIDM, projections of this person onto the individual models in the MIDM are all under the selected threshold of 3.1. As a result the MIDM does not classify any of the unknown people as belonging to the training set.
- Second, the eight people not included in the training set in step one are now included in the process of building the individuals models. Though, only images of the eight people without spectacles are included. The goal is to analyze whether or not MIDM still is capable of recognizing people as belonging to the training set when slight changes in their appearance have occurred. Here, slight changes in appearance of the eight people are modelled by the 32 synthetic images. The 32 synthetic images of the eight persons were classified correctly by the MIDM method. In Figure 13.4 III this is observed for one of the eight persons, where the projections onto the individual model of this person clearly surpass the threshold of 3.1.

Session	Shot	Clients	Impostors	
1	1	Training set (Clients)	Evaluation set (Imposters)	Test set (Imposters)
	2	Evaluation set (Clients)		
2	1	Training set (Clients)		
	2	Evaluation set (Clients)		
3	1	Training set (Clients)		
	2	Evaluation set (Clients)		
4	1	Test set (Clients)		
	2			

Table 13.2: Partitioning of data set III according to the Lausanne protocol configuration I.

In addition, the projections of the person fitted with spectacles into the individual models of the spectacle owners do not change noteworthy.

Similar results were obtained for all eight people selected to wear spectacles as seen in Figure 13.4. These results show that MIDM is a highly suitable candidate to be incorporated into a biometrical security system.

13.3 Lausanne Performance Tests

The Lausanne protocol Luetin *et al.* [39] describes how to test the XM2VTS facial database (data set III), to obtain results which can be compared to results obtained by other methods.

The Lausanne protocol uses two configurations to divide the 295 individuals of data set III into three groups consisting of 200 clients, 25 evaluation imposters and 70 testing imposters, respectively. The client group is further divided into three groups consisting of training clients, evaluation clients and testing clients, respectively. The five groups are denoted training set (training clients), evaluation set (evaluation clients + imposters) and test set (test clients + imposters) in the following. The partitioning are listed in Table 13.2 and Table 13.3 for the Lausanne protocol configuration I and II, respectively.

The performance tests are conducted by training the MIDM method using the training set. The threshold for the classifier was selected as the value at the EER for the evaluation set. The FAR and FRR of the test set were then calculated using this threshold.

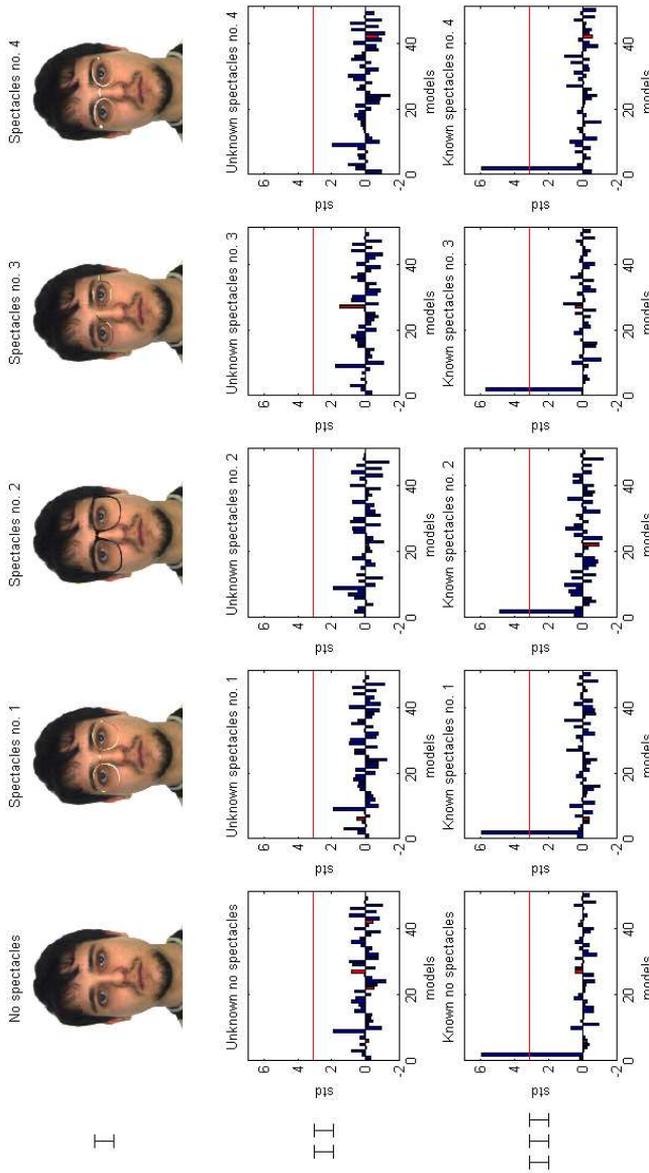


Figure 13.4: Effect of change in a person's appearance, illustrated by superimposing spectacles onto a person from data set II not wearing spectacles. (I) Original facial image without spectacles followed by four images fitted with four types of spectacles. (II) The corresponding projections into the MIDM models obtained where the person is unknown to the MIDM. (III) The corresponding projections into the MIDM models obtained where the person is known to the MIDM. The red columns in the plots (II & III) represent the model of the person from which the spectacles used to superimpose were extracted.

Session	Shot	Clients	Impostors	
1	1	Training set (Clients)	Evaluation set (Imposters)	Test set (Imposters)
	2			
2	1			
	2			
3	1	Evaluation set (Clients)		
	2			
4	1	Test set (Clients)		
	2			

Table 13.3: Partitioning of data set III according to the Lausanne protocol configuration II.

13.3.1 Participants in the Face Verification Contest, 2003

The participants in the Face Verification Contest, 2003 originated from seven institutions around the world:

- University of Surrey, UK (UniS).
- University of Kent, UK (UniK).
- The Dalle Molle Institute for Perceptual Artificial Intelligence, Switzerland (IDIAP).
- Universite Catholique de Louvain, Belgium (UCL).
- Mahanakron University of Technology, Thailand (MUT).
- Universidad Politcnica de Valencia, Spain (UPV).
- Tübitak Bilten, Turkey (TB).

In the following section results obtained from the MIDM are compared to the results obtained during the Face Verification Contest, 2003.

Method	Evaluation Set			Test Set		
	FAR	FRR	TER	FAR	FRR	TER
UniS 1	-	-	5.00	2.30	2.50	4.80
IDIAP 1	1.67	1.67	3.34	1.748	2.000	3.75
IDIAP 2	0.75	2.00	2.75	1.84	1.50	3.34
UniS 2	1.16	1.05	2.21	0.97	0.50	1.47
UCL	1.17	1.17	2.34	1.71	1.50	3.21
TB	2.34	1.00	3.34	5.61	5.75	11.36
UniS 3	0.0	0.0	0.0	0.86	0.75	1.61
UniS 4	0.33	1.33	1.36	0.48	1.00	1.48
MIDM 1	0.49	0.5	0.99	0.33	0.75	1.08
MIDM 2	0.31	0.33	0.64	0.27	0.75	1.02

Table 13.4: Error rates according to the Lausanne protocol configuration I with manual annotation of landmarks. The three highest performing methods in term of Total Error Rate (TER) are highlighted with consecutive shades of gray. Here, dark gray, medium gray and light gray indicates the highest, second highest and third highest performing method, respectively.

13.3.2 Results

The MIDM algorithm was implemented in two versions for use in testing of the **data set III** (the XM2VTS database) according to the Lausanne protocol. The two versions are denoted MIDM 1 and MIDM 2 and implemented as:

- **MIDM 1.**

The MIDM 1 version only rejects an imposter from the score of the identity the imposter uses in the attempt to gain access.

- **MIDM 2.**

The MIDM 2 version is a improvement of the MIDM 1. It rejects an imposter from the score of the identity the imposter uses in the attempt to gain access as well as rejecting the imposter if he/she scores higher in another identity of the MIDM 2 than in the one the imposter is claiming. This can be illustrated by a situation where an imposter named Bo tries to gain access using the identity Alice. The MIDM 2 evaluates Bo's score against all identities in the MIDM 2 database, and concludes that Bo resembles Bob more than Alice, i.e. the system rejects Bo because he resembles another identity of the MIDM 2 more than the one he uses in the attempt to gain access.

The results reported in Messer *et al.* [42] and results obtained from using the two versions of the MIDM algorithm, MIDM 1 and 2, are listed in Table 13.4 to Table 13.7.

Method	Evaluation Set			Test Set		
	FAR	FRR	TER	FAR	FRR	TER
IDIAP 1	1.25	1.25	2.5	1.465	2.250	3.715
IDIAP 2	0.75	0.75	1.50	1.04	0.25	1.29
TB	1.10	0.50	1.60	3.22	4.50	7.72
UniS 4	0.33	0.75	1.08	0.25	0.50	0.75
MIDM 1	0.25	0.25	0.5	0.2	0.75	0.95
MIDM 2	0.21	0.25	0.46	0.21	0.5	0.71

Table 13.5: Error rates according to the Lausanne protocol configuration II with manual annotation of landmarks. The three highest performing methods in term of Total Error Rate (TER) are highlighted with consecutive shades of gray. Here, dark gray, medium gray and light gray indicates the highest, second highest and third highest performing method, respectively.

Method	Evaluation Set			Test Set		
	FAR	FRR	TER	FAR	FRR	TER
UniS 1	-	-	14.0	5.8	7.3	13.1
IDIAP 2	1.21	2.00	3.21	1.95	2.75	4.70
UPV	1.33	1.33	2.66	1.23	2.75	3.98
UniS 4	0.82	4.16	4.98	1.36	2.5	3.86
MIDM 1	0.5	0.5	1	0.42	1.25	1.67
MIDM 2	0.34	0.5	0.84	0.33	1.25	1.58

Table 13.6: Error rates according to the Lausanne protocol configuration I with automatic annotation of landmarks (Annotations are obtained as described in Section 6.2.3). The three highest performing methods in term of Total Error Rate (TER) are highlighted with consecutive shades of gray. Here, dark gray, medium gray and light gray indicates the highest, second highest and third highest performing method, respectively.

Method	Evaluation Set			Test Set		
	FAR	FRR	TER	FAR	FRR	TER
IDIAP 2	1.25	1.20	2.45	1.35	0.75	2.10
UPV	1.75	1.75	3.50	1.55	0.75	2.30
UniS 4	0.63	2.25	2.88	1.36	2.0	3.36
MIDM 1	0.36	0.67	1.03	0.47	0.75	1.22
MIDM 2	0.28	0.5	0.78	0.28	0.75	1.03

Table 13.7: Error rates according to the Lausanne protocol configuration II with automatic annotation of landmarks (Annotations are obtained as described in Section 6.2.3). The three highest performing methods in term of Total Error Rate (TER) are highlighted with consecutive shades of gray. Here, dark gray, medium gray and light gray indicates the highest, second highest and third highest performing method, respectively.

The TER results obtained from the two MIDM methods and the all over best performing method presented in the Face Verification Contest, 2003, are taken from Table 13.4 to Table 13.7 and shown in a bar plot in Figure 13.5. The all over best performing method presented in the Face Verification Contest, 2003, was UniS 4, presented by the University of Surrey.

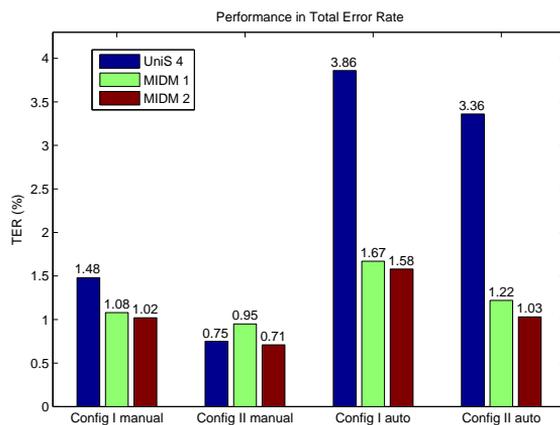


Figure 13.5: A bar plot of the TER obtained from the two MIDM methods and from UniS 4.

The results will be discussed in the following section.

13.4 Discussion

From the results obtained in the identification test in Section 13.2.1 it is clear that the best performance of MIDM is obtained when both the geometrical and the photometrical information are used. The MIDM algorithm has a higher rate of performance than the Fisherface method (Table 13.1) and has the following advantages:

- MIDM is very intuitive. The process used in MIDM to determine whether or not a person is classified as belonging to a specific model of the MIDM is a one-dimensional problem.
- MIDM is highly flexible and changes in one model do not interfere with the other models of the MIDM.
- The scalability of MIDM is very high, since the MIDM method can be easily parallelized to use in a cluster of computers.

- No training is needed for the classifier to estimate the FAR. (Estimating FRR requires training).

In Section 13.2.3 a verification test (a 25-fold cross-validation of **data set II**) was performed in both a best and a worst case scenario yielding Equal Error Rates as low as 0.3 and 1.8%, respectively. This is a very satisfying result obtained from an image data set as large as 700 images.

From the robustness test in Section 13.2.4 a satisfying result was obtained. Changing a persons appearance (i.e. whether or not a person wears spectacles as well as change in the appearance of the spectacles.) does not result in change of identity. This result points to the conclusion that face databases should be captured omitting spectacles.

The performance test in Section 13.3 showed that the MIDM algorithm is superior to all the participating methods in the Face Verification Contest in 2003 (Figure 13.5).

The MIDM algorithm proposed in Chapter 11 satisfies the demands settled at the beginning of this thesis, i.e. the objective was to design and develop a robust facial recognition algorithm constructed in a simple and easy adaptable setup. The results presented in this chapter shows that the MIDM algorithm is a robust and superior face recognition algorithm that is a highly qualified candidate to be used in a facial recognition application.

Part IV

Implementation

Implementation

14.1 Overview

This chapter describes the different implementations developed during this thesis. These are:

- **FaceRec**, a Delphi 7 implementation of an automatic facial recognition process using AAM and MIDM.
- A DLL of selected functions of the AAM-API [51].
- A small C++ program used to collect the shape free images and save these as texture vectors in a Matlab format.
- A Matlab function used to construct the MIDM Model File.
- Various Matlab functions of the described topics in this thesis.

A CD-ROM is enclosed containing all the above mentioned implementations, source code and digital versions of this thesis.

The countless Matlab scripts used to generate the statistical results and the 3D reconstruction are not included. The Matlab scripts are very comprehensive

and left out, since they are very machine specific. The specific content of the CD-ROM is listed in Appendix F.

14.2 FaceRec

FaceRec is a Delphi 7 implementation of a real time automatic facial recognition system. FaceRec is implemented for the face recognition task (one to many search) and is not expected to encounter unknown persons. FaceRec uses AAM for face detection and MIDM for face recognition. However, due to time constraints only texture information is used for face recognition by the MIDM. This abolishes the need to implement solution algorithms to Procrustes analysis as well as the eigenvalue problem in Delphi. A screen shot of FaceRec is displayed in Figure 14.1. A quick user guide is provided in Appendix D.

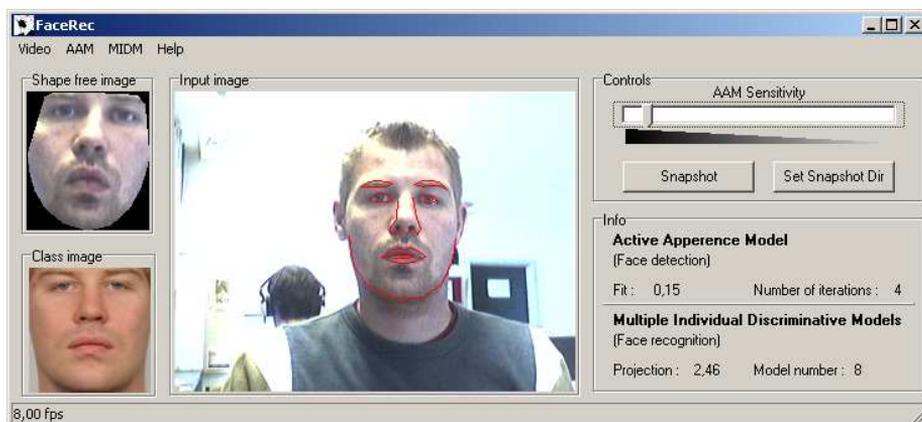


Figure 14.1: The main window of the FaceRec application, which contains an input from a webcam, the shape free image of a captured face, the class image of the recognized person, a control panel and an AAM/MIDM information panel.

A weakness of MIDM is that it is only trained to handle facial images and the algorithm assumes that the face detection only provides facial images. This causes a problem when MIDM is introduced to non-facial or half facial images. A snapshot of a false recognition due to a non-facial image is displayed in Figure 14.2. It is recommended that the sensitivity of the AAM is kept so high that only faces are captured. A AAM Model File and a MIDM Model File built using the IMM Frontal Face Database (**data set I**) is included on the CD-ROM, however face detection (AAM) will only be optimal when used in the same lighting condition as the pictures used to build the AAM Model File.



Figure 14.2: The FaceRec application obtaining a false recognition due to a non-facial image.

14.2.1 FaceRec Requirements

The FaceRec application is designed for the windows platform and is fully tested on a Windows-XP machine. FaceRec requires a webcam for image acquisition and is compatible with all webcam drivers recognized by windows. For face detection an AAM Model File (.amf) must be provided, which can be constructed by the AAM-API. For face recognition a MIDM Model File (.mmf) must be provided, which can be constructed by the `makeMIDModel.m` Matlab function included on the inclosed CD-ROM.

14.2.2 AAM-API DLL

A DLL was developed enabling usage of selected classes/functions of the AAM-API C++ implementation in a Delphi 7 application. The `aam_api.dll` and source code are included on the enclosed CD-ROM. A table of the included functions and a short description are provided in Appendix E. For more information please consult the source code and the AAM-API documentation. Notice that this DLL incorporates the AAM-API color model, so when building an AAM Model File be sure to use the color option.

14.2.3 Make MIDM Model File

The Matlab function `makeMIDModel.m` is used to construct the MIDM Model File used by `FaceRec`. The function only implements the MIDM algorithm for texture information. In order to retrieve texture vectors from free shape images a small C++ application is developed. For more information please consult the source code or type “help makeMIDModel” in Matlab.

14.2.3.1 Get Texture Vectors

The `getTextureVectors.exe` C++ application opens an AAM Model File and saves the texture vectors of the free shape images obtained from all images residing in the same folder as the AAM Model File, in a Matlab format.

14.3 Matlab Functions

The following Matlab functions were developed during this thesis and are included on the enclosed CD-ROM for possible future use.

- `flda.m`
Fisher linear discriminant analysis, described in Section 9.2.
- `histEqImage.m`
Histogram equalization of an image, described in Section 8.1.1.
- `kNearNeighbor.m`
 k nearest neighbor algorithm.
- `lightCompensationModels.m`
Removal of specific illumination conditions based on illumination compensation models, as described in Section 8.1.2.
- `lpp.m`
Locality preserving projections, described in Section 9.3.
- `pca_minMem.m`
Principal component analysis, using the method described in Section 9.1.2, to achieve a speedup and reduction in memory usage.

- `pca_numberOfEigenvectors.m`
This function calculates the number of eigenvectors needed to capture a certain percentage of the image variance, as described in Eq. 7.2.

`pca_minMem.m` and `pca_numberOfEigenvectors.m` are used by `makeMIDModel.m`. The `readasf.m` Matlab function distributed with the AAM-API for reading `asf` files is also included in the inclosed CD-ROM. The results in Chapter 10 for the KFDA method was obtained by the Generalized Discriminant Analysis GDA Matlab package provided by Baudat *et al.* [3].

14.4 Pitfalls

Two tricky issues of implementation are highlighted in the following, passing arrays and the Matlab `eig` function.

14.4.1 Passing Arrays

When passing an array from Delphi to C++ the array is received by C++ as a array pointer and the length of the array. An example is provided in the following by the function in the `aam_api.dll` for retrieving the center of gravity of a shape:

Call from Delphi:

```
procedure shapeCOG_dll(shape:CAAMShapeHandle;  
                      var shapeCOG : array of double); stdcall;
```

Function in the DLL (C++):

```
void __declspec(dllexport) shapeCOG(CAAMShapeHandle shape,  
                                   double *shapeCOG, int arraylength);
```

14.4.2 The Matlab Eig Function

When solving the eigenvalue problem in Matlab by using the function `eig` one has to consider that Matlab release 14 (used in this thesis) does not guarantee

that eigenvalues and eigenvectors are sorted in ascending order (which was the case for Matlab release 13). To overcome this the following code snippet can be used:

```
[eigV eigD]=eig(covarianceMatrix);  
%Most important eigenvalue at buttom (right) Needs  
%to be sorted, there is a "bug" (change) in Matlab R 14 eig.  
[sorted I] = sort(sum(eigD));  
eigV = eigV(:,I);  
eigD = diag(sorted);
```

Part V

Discussion

Future Work

The following sections describe interesting areas which could improve the facial recognition of the MIDM algorithm.

15.1 Light Normalization

One way to improve the robustness and flexibility of the MIDM algorithm is to improve light normalization/correlation in the preprocess step. Looking into methods that can remove specific light sources leaving the facial image in a diffuse lighting is very interesting and can lead to the following improvements:

- Successful implementation of the MIDM algorithm in environments with large variation in lighting conditions.
- An increase in performance of the MIDM algorithm, since it does not need to be trained to be capable of handling multi-lighting conditions.
- The enrollment of new persons in the MIDM scheme would be a much simpler and more robust process.

15.2 Face Detection

From an implementation point of view proper face detection is very important in a face recognition process. The performance of a face recognition scheme can be notably improved by minimizing the false detection of faces. As described in Chapter 14 the MIDM algorithm shows a weakness when it is presented to a non-facial image, since it is not trained to handle the variation of non-facial images, resulting in a possible misclassification. In addition, the face detection algorithm (AAM) used in the implementation described in Chapter 14 is very sensitive to specific lighting condition, which causes a problem. Decreasing the AAM sensitivity towards specific lighting conditions would be an area of great interest.

15.3 3D Facial Reconstruction

As described in Chapter 12 facial reconstruction in 3D from 2D images is a very interesting area. The end goal is extraction of more discriminative 3D information from which a better face recognition performance could be obtained compared to that from 2D images. Reconstruction and recognition of not frontal face images through 3D images are of great interest as well. The interest in this area is not limited to facial recognition, since reconstruction of a three-dimensional shape without use of 3D scanners would be useable in many applications as well.

CHAPTER 16

Discussion

16.1 Summary of Main Contributions

The following four sections describe the main contributions of this thesis to facial recognition.

16.1.1 IMM Frontal Face Database

As a part of this thesis the IMM Frontal Face Database was constructed for use in the initial testing of feature extraction methods. The database consists of 120 images of 12 persons and is described in detail in Appendix A. The database only contains small variation (in term of facial poses, illumination conditions, etc.), which makes it an ideal database when working in a low-dimensional feature space, e.g. in dimension one to five. However, to obtain reliable performance tests one need to use other databases with more variation.

16.1.2 The MIDM Algorithm

A central part in this thesis is the development of the Multiple Individual Discriminative Models (MIDM) algorithm, as described in Chapter 11. MIDM is a feature extraction and identity matching algorithm, which has been proven very robust and adaptive (see Chapter 13).

16.1.3 A Delphi Implementation

A Delphi implementation of an automated face recognition scheme was constructed as described in Chapter 14. The implementation includes a DLL, incorporating a C++ implementation of Active Appearance Models (AAM-API [51]). The implementation is designed for the windows platform and requires a webcam for image acquisition.

16.1.4 Matlab Functions

The following Matlab functions were constructed and used in this thesis:

- `flda.m`
Fisher linear discriminant analysis, described in Section 9.2.
- `histEqImage.m`
Histogram equalization of an image, described in Section 8.1.1.
- `kNearNeighbor.m`
 k nearest neighbor algorithm.
- `lightCompensationModels.m`
Removal of specific illumination conditions based on illumination compensation models, as described in Section 8.1.2.
- `lpp.m`
Locality preserving projections, described in Section 9.3.
- `pca_minMem.m`
Principal component analysis, using the method described in Section 9.1.2, to achieve a speedup and reduction in memory usage.

16.2 Conclusion

The increase in complexity in consumer goods such as automobiles and home appliances provides for a new market for technologies such as face recognition. Additionally, the recent events in the world has spurred a public demand for security and safety in public places; demands that could be partially met by deployment of these technologies.

The four objectives of this thesis were: To discuss and summarize the process of facial recognition, to look at currently available facial recognition techniques, to design and develop a robust facial recognition algorithm and finally an implementation of this new algorithm.

In Chapter 2 to Chapter 10 this thesis presents a comprehensive overview of the area of facial recognition, satisfying the two first objectives. The third objective of this thesis is satisfied by the work presented in Chapter 11 to Chapter 13 by the design and development of MIDM, a new feature extraction and identity matching algorithm. By tests in the mentioned chapters MIDM has been proven superior to all participating methods in the Face Verification Contest at The Audio Video Biometric Person Authentication Conference in 2003. MIDM is to participate in the Face Verification Contest at the International Conference on Biometrics in Hong Kong in January, 2006.

The last objective is satisfied by the work presented in Chapter 14. The implementation of the MIDM algorithm does not require any special hardware besides a webcam for collection of input images.

Though MIDM has been developed to be used in the process of facial recognition the algorithm can easily be adapted to recognize other objects. This makes MIDM a useful algorithm in more than one sense.

With the completion of this thesis, an important step in addressing the quality and reliability of face recognition schemes has been completed.

List of Figures

1.1	Comparison of machine readable travel documents (MRTD) compatibility with six biometric techniques.	2
1.2	“The Wife and the Mother-in-Law” by Edwin G. Boring.	3
3.1	Relation of False Acceptance Rate (FAR), False Rejection Rate (FRR) with the distribution of clients, imposters in a verification scheme.	15
3.2	The Correct Identification Rates (CIR) plotted as a function of gallery size.	18
3.3	The average Correct Identification Rates (CIR) of the three highest performing systems in the Face Recognition Vendor Test 2002 from a age perspective.	18
3.4	The average Correct Identification Rates (CIR) of the three highest performing systems in the Face Recognition Vendor Test 2002 from a time perspective.	19
4.1	The four general steps in facial recognition.	22
5.1	Small interpersonal variations illustrated by identical twins.	26

5.2	A sample of 33 facial images produced from one original image.	28
5.3	Results of the exploration of facial submanifolds.	29
5.4	An illustration of the problem of not being capable of satisfactory approximating the manifold when only having a small number of samples.	31
6.1	An example of ten images of one person from the IMM Frontal Face Database.	37
6.2	The 73-landmark annotation scheme used on the IMM Frontal Face Database.	38
6.3	Examples of 14 images of one female and one male obtained from the AR database.	39
6.4	The 22-landmark annotation scheme used on the AR database.	40
6.5	Examples of 8 images of one female and one male obtained from the XM2VTS database.	42
6.6	The 64-landmark annotation scheme used on the XM2VTS database.	43
6.7	Five samples from 3D Face Database constructed in [49].	44
7.1	Illustration of the subsampling of an image in a pyramid fashion. Which enables the capture of different size of faces.	50
7.2	Example of 10 Eigenfaces.	51
7.3	Diagram of the neural network developed by Rowley <i>et al.</i> [47].	53
7.4	Diagram displaying an improved version of the neural network in Figure 7.3.	54
7.5	Example of full Procrustes analysis.	55
7.6	Face detection (approximations) obtained by AAM, when the model is initialized close to the face.	57

8.1	Examples of histogram equalization used upon two images to obtain standardized images with maximum entropy.	60
8.2	Removal of specific illumination conditions from facial images.	62
8.3	A close-up of the faces reconstructed in Figure 8.2.	62
8.4	Illumination compensation maps used for removal of specific illumination conditions.	63
9.1	An example of PCA in two dimensions.	66
9.2	An example of FLDA in two dimensions.	69
10.1	The PCA scatter plot matrix of the combined features from data set I	81
10.2	The LPP scatter plot matrix of the combined features from data set I	82
10.3	The FLDA scatter plot matrix of the combined features from data set I	83
10.4	The KFDA scatter plot matrix of the combined features from data set I	84
10.5	False identification rates obtained from the 50/50 test, where data set I is divided into two equal sized sets, a training set and a test set.	85
10.6	The false identification rates obtained from the ten-fold cross-validation test.	86
11.1	Overview of the MIDM algorithm.	93
11.2	Example of a Delaunay triangulation of the landmark scheme in data set II	94
12.1	Process of projecting 3D landmarks into a 2D image, which is used to calculate a fit of the 3D shape.	100

12.2	Example of retrieved depth information from a 2D frontal face image.	101
12.3	Example of retrieved depth information from a 2D frontal face image and a 2D profile face image.	102
13.1	Importance of geometrical information.	108
13.2	The 10%, 15% and 25% pixels weights of highest value most important for discrimination between the test persons of data set II	109
13.3	False Acceptance Rate and False Rejections Rate graphs and Receiver Operating Characteristic (ROC) curve obtained by the 25-fold cross-validation for the best and worst case scenario, respectively.	111
13.4	Effect of change in a persons appearance, illustrated by superimposing spectacles onto a person from data set II not wearing spectacles.	114
13.5	A bar plot of the TER obtained from the two MIDM methods and from UniS 4.	118
14.1	The main window of the FaceRec application.	124
14.2	The FaceRec application obtaining a false recognition due to a non-facial image.	125
D.1	The main window of the FaceRec application.	174
D.2	The video driver list popup box.	174
D.3	The specific “Video Source” panel included in the drivers of the Philips PCVC690K webcam.	175
D.4	The FaceRec application capturing a face and recognizing it as model (person) number 8.	176
F.1	The contents of the enclosed CD-ROM.	182

List of Tables

3.1	Participants in the Face Recognition Vendor Test 2002.	16
3.2	The characteristics of the highest performing systems in the Face Recognition Vendor Test 2002.	17
7.1	Categorization of methods for face detection within a single image.	49
9.1	Dimensionality reduction methods.	65
13.1	Correct identification rates.	108
13.2	Partitioning of data set III according to the Lausanne protocol configuration I.	113
13.3	Partitioning of data set III according to the Lausanne protocol configuration II.	115
13.4	Error rates according to the Lausanne protocol configuration I with manual annotation of landmarks.	116
13.5	Error rates according to the Lausanne protocol configuration II with manual annotation of landmarks.	117

13.6	Error rates according to the Lausanne protocol configuration I with automatic annotation of landmarks.	117
13.7	Error rates according to the Lausanne protocol configuration II with automatic annotation of landmarks.	117
E.1	Function list over the <code>aam_api.dll</code>	178
E.2	Function list over the <code>aam_api.dll</code> continued.	179

APPENDIX A

The IMM Frontal Face Database

The IMM Frontal Face Database

An Annotated Dataset of 120 Frontal Face Images

Jens Fagertun and Mikkel B. Stegmann
Informatics and Mathematical Modelling, Technical University of Denmark
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark

29. August 2005

Abstract

This note describes a data set consisting of 120 annotated monocular images of 12 different frontal human faces. Points of correspondence are placed on each image so the data set can be readily used for building statistical models of shape. Format specifications and terms of use are also given in this note.

Keywords: Annotated image data set, frontal face images, statistical models of shape.

1 Data Set Description

This database consists of 12 people (all male). A total of 10 frontal face photos has been recorded of each person. The data set is containing different facial poses captured over a short period of time, with a minimum of variance in lighting, camera position, etc.

All photos are annotated with landmarks defining the eyebrows, eyes, nose, mouth and jaw, see Figure 1. The annotation of each photo is stored in the ASF format, described in Appendix A.

2 Specifications

2.1 General Specifications

2.1.1 Specifications of Test Persons

All test persons are males, not wearing glasses, hats or other accessories.

2.1.2 Specifications of Facial Expressions

Table 1 lists the facial expressions captured in this data set.

Facial expressions	Description
No expression	The normal facial pose
Relaxed happy	Smiling vaguely (lips closed)
Relaxed thinking	The facial expression is a little tense (try to multiply 57*9 ;c)

Table 1: Specifications of facial expressions.

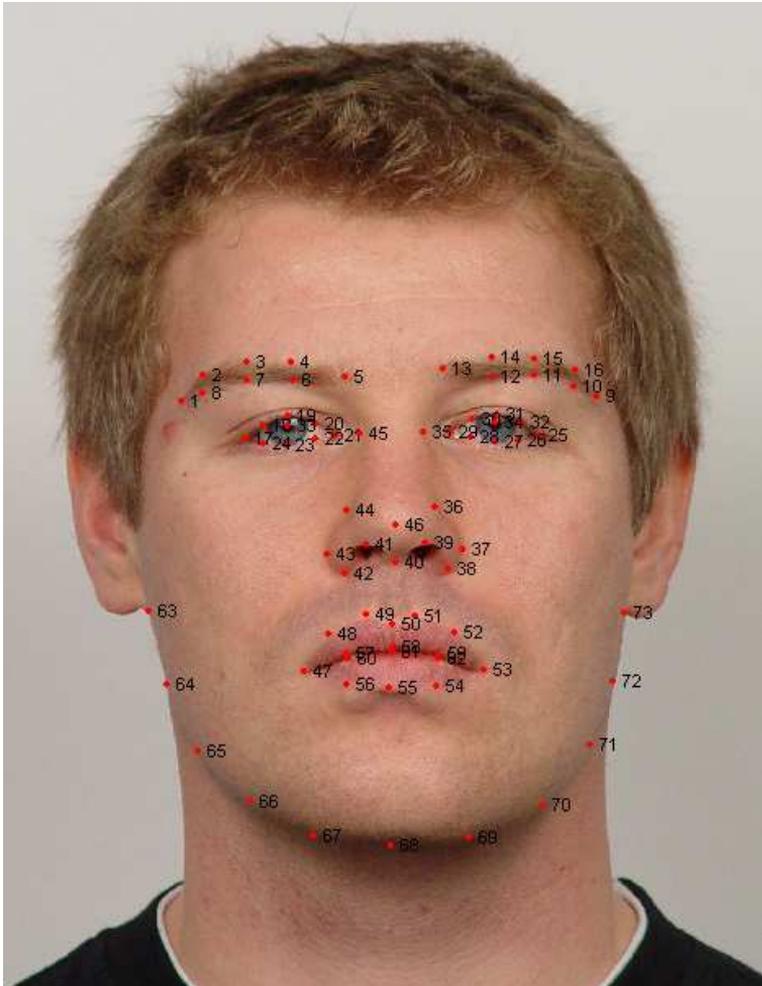


Figure 1: The 73 landmarks annotation defining the facial features; eye-brows, eyes, nose, mouth and jaw.

2.1.3 Image Format and Naming

The images are in JPEG format and named `XX_YY.jpg` where `XX` is the person id and `YY` the photo number. Table 2 shows the correspondence between the photo number and the facial expression.

Photo number	Facial expression
01 to 06	No expression
07 to 08	Relaxed happy
09 to 10	Relaxed thinking

Table 2: Photo number specification.

2.1.4 Annotation Specifications

All photos were annotated with 73 landmarks. Table 3 specifies the correspondence between annotation landmarks and facial features. For the precise landmark placements see Figure 1.

Facial Features	Annotation landmarks
Right eyebrow	1 - 8
Left eyebrow	9 - 16
Right eye	17 - 24
Left eye	25 - 32
Right eyes pupil	33
Left eyes pupil	34
Nose	35 - 46
Mouth	47 - 62
Jaw	63 - 73

Table 3: The 73 landmarks annotation defining the facial features; eyebrows, eyes, nose, mouth and jaw.

2.2 Studio Specifications

Figure 2 displays the studio setup.

2.2.1 Specification of Backdrop

In this data set a white projector screen is used as backdrop, which is a uniform nonreflecting surface, distinguishable from the test persons skin, hair and clothes. The camera lens has to be parallel to the background.

2.2.2 Specifications of Camera and Person Placement

The person was sitting down on an office chair and filmed with a straight back. The camera was placed in the same height as the test person eyes. The face of the test person was parallel to the background and the camera lens. An example of a full size image is shown in Figure 3. The camera captures additional space above and below the head of the test person in order to insure, that all test persons can be recorded without altering the studio setup.

2.2.3 Specifications of Light

The diffuse light is coming solely from two spot lights. The light was bounced off using white umbrellas. In the studio there was no interference from sun light, room light etc.

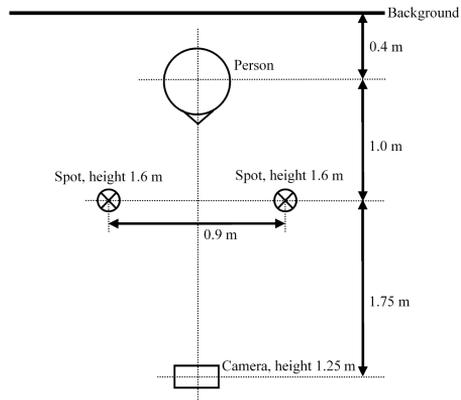


Figure 2: Studio setup: Height of the spots is from floor to bulb. Height of the camera is from floor to center of lens.



Figure 3: Example of a full size image.

2.3 Technical Specifications

2.3.1 Camera Specifications

Name	Sony - Cyber-shot
Image size	5.0 megapixels
Model no	DSC-F717

2.3.2 Image Specifications

Focal length	48 mm ¹
Compression	Fine/Normal (jpg) ²
Iso speed	200
Aperture	F/2.4
Shutter speed	1/100 s
Image size	2560 × 1920 pixels
Exposure Compensation	+ 0.7 EV
White balance	Custom ³

2.3.3 Spot Light Specifications

Spot Light:

Manufacturer	Hedler
Name	Videolux 300

Bulb:

Manufacturer	Osram
Type	Halogen
Name	Photo Optic Lamp
No	64515
Watt	300 w

3 The Work Process Protocol of One Test Person

1. The camera position is placed accordingly to the specifications above (approx 1.25 m from floor to lens and same level as the eyes of the test person).
2. The test person is explained the meaning of the pose to be recorded, in terms of the facial expression.
3. The test person is distracted for one second to "reset" facial features (ex. rolls with the head), and assumes the wanted pose, the photo is recorded. (Make sure the test persons pitch, roll and yaw is not to critical compared to the camera lens. Nostrils should be just visible).
4. Item 3. is repeated until all the photos of this pose are recorded.
5. Item 2. is repeated for all the wanted poses.
6. Record the test persons age.

¹Equivalent to 190 mm with a 35 mm FOV.

²The average file size was 1.9MB.

³The white balance on the camera was calibrated to the backdrop with the spotlights.

4 Things to Improve

- Green background (not shiny).
- More diffuse lighting (the lighting can be placed further away from the test person).
- A real photo chair (not an office chair).
- Fixed focal length (a camera where you can see the current exact focal length).
- Strict rules for test persons pitch, roll and yaw.

5 Terms of Use

The data set can be freely used for education and research. The only requirement is that a reference to this note is given.

A ASF – AAM Shape Format Specification

An ASF file is structured as a set of lines separated by a CR character. Anywhere in the file, comments can be added by starting a line with the '#' character. Comment lines and empty lines are discarded prior to parsing. The layout of an ASF file is as follows:

- Line 1 contains the total number of points, n , in the shape.
- Line 2 to $n+1$ contains the point information (one line per point) such as the point location, type, connectivity etc., see below. Hence, quick and simple access is preferred over data compactness.
- Line $n+2$ contains the *host image*, i.e. the filename of the image where the annotation is defined.

The formal point definition is:

```
point := <path#> <type> <x-pos> <y-pos> <point#> <connects from> <connects to>
```

<path#> The path that the point belongs to. Points from different paths must not be interchanged (in the line order).

<type> A bitmapped field that defines the type of point:

- Bit 1: Outer edge point/Inside point
- Bit 2: Original annotated point/Artificial point
- Bit 3: Closed path point/Open path point
- Bit 4: Non-hole/Hole point

Remaining bits should be set to zero. An inside artificial point which is a part of an closed hole, has thus the type: $(1 \ll 1) + (1 \ll 2) + (1 \ll 4) = 1 + 2 + 4 = 7$.

<x-pos> The relative x-position of the point. Obtained by dividing image coordinates in the range $[0; \text{image width}-1]$ by the image width (due to strange historic reasons...). Thus, pixel $x = 47$ (the 48th pixel) in a 256 pixel wide image has the relative position $47/256 = 0.18359375$.

<y-pos> The relative y-position of the point. Obtained by dividing image coordinates in the range $[0; \text{image height}-1]$ by the image height (due to strange historic reasons...). Thus, pixel $y = 47$ (the 48th pixel) in a 256 pixel tall image has the relative position $47/256 = 0.18359375$.

<point#> The point number. First point is zero. This is merely a service to the human reader since the line at where the point occurs implicitly gives the real point number.

<connects from> The previous point on this path. If none <connects from> == <point#> can be used.

<connects to> The next point on this path. If none <connects to> == <point#> can be used.

Further, the following format rules apply:

- Fields in a point specification are separated by spaces or tabs.
- Path points are assumed to be defined clockwise. That is; the outside normal is defined to be on left of the point in the clockwise direction. Holes are thus defined counter-clockwise.
- Points are defined in the fourth quadrant. Hence, the upper left corner pixel is (0,0).
- Isolated points are signaled using <connects from> == <connects to> == <point#>.
- A shape must have at least one outer edge. If the outer edge is open, the convex hull should determine the interior of the shape.

Example ASF file

```
<BOF>
#####
#
#   AAM Shape File - written: Monday May 08 - 2000 [15:22]
#
#####

#
# number of model points
#
83

#
# model points
#
# format: <path#> <type> <x rel.> <y rel.> <point#> <connects from> <connects to>
#
0 0 0.07690192 0.44500541 0 82 1
0 0 0.09916206 0.42914406 1 0 2
0 0 0.12925033 0.39573063 2 1 3

...

0 0 0.07579006 0.52910086 80 79 81
0 0 0.06128729 0.49762829 81 80 82
0 0 0.05858913 0.46610570 82 81 0

#
# host image
#
F1011flb.bmp

<EOF>
```

APPENDIX B

A face recognition algorithm based on MIDM

A face recognition algorithm based on multiple individual discriminative models

Jens Fagertun, David Delgado Gomez, Bjarne K. Ersbøll, Rasmus Larsen

Abstract—In this paper, a novel algorithm for facial recognition is proposed. The technique combines the color texture and geometrical configuration provided by face images. Landmarks and pixel intensities are used by Principal Component Analysis and Fisher Linear Discriminant Analysis to associate a one dimensional projection to each person belonging to a reference data set. Each of these projections discriminates the associated person with respect to all other people in the data set. These projections combined with a proposed classification algorithm are able to statistically deciding if a new facial image corresponds to a person in the database. Each projection is also able to visualizing the most discriminative facial features of the person associated to the projection. The performance of the proposed method is tested in two experiments. Results point out the proposed technique as an accurate and robust tool for facial identification and unknown detection.

Index Terms—Face recognition, Principal Component Analysis, Fisher Linear Discriminant Analysis, Biometrics, Multi-Subspace Method.

I. INTRODUCTION

Regrettable events which happened during the last years (New York, Madrid) have revealed flaws in the existing security systems. The vulnerability of most of the current security and personal identification system is frequently shown. Falsification of identity cards or intrusion into physical and virtual areas by cracking alphanumeric passwords appear frequently in the media. These facts have triggered a real necessity for reliable, user-friendly and widely acceptable control mechanisms for person identification and verification.

Biometrics, which bases the person authentication on the intrinsic aspects of a human being, appears as a viable alternative to more traditional approaches (such as PIN codes or passwords). Among the oldest biometrics techniques, fingerprint recognition can be found. It is known that this technique was used in China around 700 AD to officially certify contracts. Afterwards, in Europe, it was used as person identification in the middle of the 19th century. A more recent biometric technique used for people identification is iris recognition [8]. It has been calculated that the chance of finding two randomly formed identical irises is one in 10^{78} (The population of the earth is below 10^{10}) [7]. This technique has started to be used as an alternative to passport in some airports in United Kingdom, Canada and Netherlands. It is also used as employee control access to restricted areas in Canadian airports and in the New York JFK airport. The inconvenient of these techniques is the necessity of interaction with the individual who wants to be identified or authenticated. This fact has caused that face recognition, a non-intrusive technique, has

increased received the interest from the scientific community in recent years.

The first developed techniques that aimed at identifying people from facial images were based on geometrical information. Relative distances between key points, such as mouth corners or eyes, were calculated and used to characterize faces [17]. Therefore, most of the developed techniques during the first stages of facial recognition focused on the automatic detection of individual facial features. However, facial feature detection and measurements techniques developed to date are not reliable enough for the geometric feature based recognition, and such geometric properties alone are inadequate for face recognition because rich information contained in the facial texture or appearance is discarded [6], [13]. This fact produced that gradually most of the geometrical approaches were abandoned for color based techniques, which provided better results. These methods aligned the different faces to obtain a correspondence between pixels intensities. A nearest neighbor classifier used these aligned values to classify the different faces. This coarse method was notably enhanced with the appearance of the Eigenfaces technique [15]. Instead of directly comparing the pixel intensities of the different face images, the dimension of these input intensities were first reduced by a principal component analysis (PCA). This technique settled the basis of many of the current image based facial recognition schemes. Among these current techniques, Fisherfaces can be found. This technique, widely used and referred [2], [4], combines the Eigenfaces with Fisher linear discriminant analysis (FLDA) to obtain a better separation of the individuals. In Fisherfaces, the dimension of the input intensity vectors is reduced by PCA and then FLDA is applied to obtain a good separation of the different persons.

After Fisherfaces, many related techniques have been proposed. These new techniques aim at providing a projection that attain a good person discrimination and also are robust at differences in illumination or image pose. Kernel Fisherfaces [16], Laplacianfaces [10] or discriminative common vectors [3] can be found among these new approaches. Typically, these techniques have been tested assuming that the image to be classified corresponds to one of the people in the database. In these approaches, the image is usually classified to the person with the smallest Euclidean distance.

However, some inconveniences appear when the person to be analyzed may not belong to the data set. In this case, a criterium to decide if the person belongs to the data set has to be chosen. E.g. only people with an euclidian distance less than a given threshold are considered as belonging to the data set. However, this threshold has not to be necessarily the same

for all the classes (different persons) and different thresholds would need to be found. The estimation of these thresholds is not straightforward and additional data might be needed.

In this work, a new technique that addresses the different inconveniences is proposed. The proposed technique takes advantage of two novelties in order to deal with these inconveniences. First, not only the texture intensities are taken into account but also the geometrical information. Second, the data are projected into n one-dimensional spaces instead of a $(n - 1)$ -dimensional space, where n is the number of people in the data set.

Each of these individual models aims at characterizing a given person uniquely. This means that every person in the data set is represented by one model. These multi one-dimensional models allow to statistically interpret the "degree of membership" of a person to the data set and to detect unknowns. Furthermore, these two facts have several advantages in interpretability, characterization, accuracy and easiness to update the model.

II. ALGORITHM DESCRIPTION

The proposed algorithm is made up of two steps. In the first step, an individual model is built for each person in the database using the color and geometrical information provided by the available images. Each model characterizes a given person and discriminates it from the other people in the database. The second step carries out the identification. A classifier, related with the standard Gaussian distribution, decides if a face image belongs to one person in the database or not. In this section, the two parts of the algorithm are described in detail. A diagram of the algorithm is displayed in Fig. 1. This diagram will be referred during the description of the algorithm to obtain an easier understanding.

A. Creating the individual models

1) *Obtaining the geometry of the face:* The geometrical characterization of a given face is obtained by means of the theory of statistical shape analysis [1]. In this theory, objects (faces) are represented by shapes. According to Kendall [11], a shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object. In order to describe a shape, a set of landmarks or points of correspondence that matches between and within populations are placed on each face. As an example, Fig. 2A displays a set of 22 landmarks. These landmarks indicate the position of the eyebrows, eyes, nose, mouth, jaw and size of a given face.

To obtain a shape representation according to the definition, the obtained landmarks are aligned in order to remove the location, rotational and scaling effects. To achieve this goal, the 2D-full Procrustes analysis is carried out. Briefly, let:

$$\mathbf{X} = \{\mathbf{x}_i\} = \{x_i + i \cdot y_i\}, \quad i = 1, \dots, n$$

be a set of n landmarks expressed in complex notation. In order to apply full Procrustes analysis, the shapes are initially

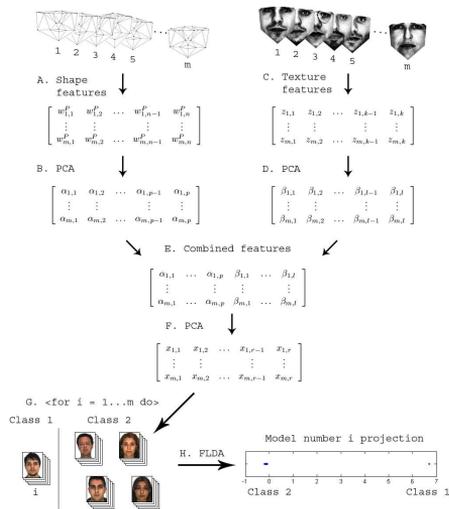


Fig. 1. Algorithm overview. A: Landmarks alignment using full Procrustes analysis. B: PCA on aligned landmarks to remove redundancy. C: Texture normalization using global histogram equalization. D: PCA on normalized texture to remove redundancy. E: Combining shape and texture features. F: PCA on combined features to remove redundancy. G & H: In turn build the individual model using FLDA.

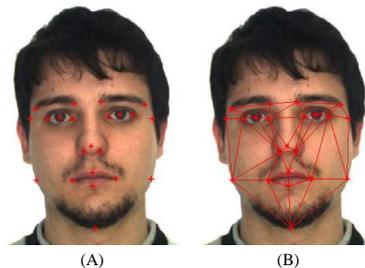


Fig. 2. (A) Set of 22 landmarks placed on a face image. (B) The Delaunay triangulation of the 22 landmarks.

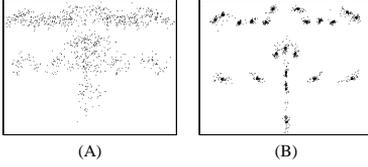


Fig. 3. (A) Superimposition of the sets of 22 landmarks obtained over 49 different face images. (B) Alignment of the landmarks.

centered. To center the different shapes, the mean of the shape, \bar{x} , is subtracted from each landmark:

$$w_i = x_i - \bar{x}, \quad i = 1, \dots, n$$

The full Procrustes mean shape [12], $\hat{\mu}$, is found as the eigenvector corresponding to the largest eigenvalue of the complex sum of squares and products matrix

$$\sum_{i=1}^n w_i w_i^* / (w_i^* w_i)$$

where w_i^* denotes the transpose of the complex conjugate of w_i . Using this Procrustes mean shape, the full Procrustes coordinates of w_1, \dots, w_n (Fig. 1 A) are obtained by

$$w_i^P = w_i^* \hat{\mu} w_i / (w_i^* w_i) \quad i = 1, \dots, n$$

Fig. 3A displays the superimposition of the set of 22 landmarks described in Fig. 2, obtained on 49 different face images. The result obtained after applying the full Procrustes alignment on these landmarks can be observed in Fig. 3B. In order to remove redundancy in the data, a Principal Component Analysis is applied to the aligned landmarks (Fig. 1 B).

2) *Texture formulation*: To form a complete model of the face appearance, the algorithm also captures the texture information provided by the pixels. In order to collect this texture representation, the Delaunay triangulation of every shape is obtained. The Delaunay triangulation connects the aligned landmark set of each image by a mesh of triangles, so no triangle has any of the other points of the set inside its circumcircle. The Delaunay triangulation obtained for each image is warped onto the Delaunay triangulation of the mean shape. The Delaunay triangulation of the 22 landmarks is displayed in Fig. 2B.

Formally, let I be a given image and M the mean shape previously obtained. Let $u_1 = [x_1, y_1]$, $u_2 = [x_2, y_2]$ and $u_3 = [x_3, y_3]$ denote the vertices of a triangle T in I , and let v_1, v_2 and v_3 be the associated vertices of the corresponding triangle in M . Given any internal point $\hat{u} = [x, y]$ in the triangle T , the corresponding point in the associated triangle in the mean shape can be written as $\hat{v} = \alpha v_1 + \beta v_2 + \gamma v_3$ where:

$$\begin{aligned} \alpha &= 1 - (\beta + \gamma) \\ \beta &= \frac{yx_3 - x_1y - x_3y_1 - y_3x + x_1y_3 + xy_1}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2} \\ \gamma &= \frac{xy_2 - xy_1 - x_1y_2 - x_2y + x_2y_1 + x_1y}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2} \end{aligned}$$

This transformation extracts the texture of a given face image. A histogram equalization is applied to the collected texture to reduce the effects of differences in illumination [9]. This histogram equalization is performed independently in each of the three color channels. Afterwards, the three color channels are converted into gray scale to obtain a more compact representation (Fig. 1 C).

Similarly to the shape analysis, a PCA is conducted in the texture data to reduce dimensionality and data redundancy (Fig. 1 D). However, notice that the large dimension of the texture vectors will produce memory problems because of the huge dimension of the covariance matrix. In order to avoid this difficulty, the Eckart-Young theorem is used [5]. Formally, let D represents the texture data matrix composed by s n -dimensional texture vectors after the mean of the texture vectors has been subtracted from each one of them ($s < n$). Then the $n \times n$ dimensional covariance matrix can be written as:

$$\Sigma_D = \frac{1}{s} DD^T$$

Let Σ_S be the smaller $s \times s$ dimensional matrix defined by

$$\Sigma_S = \frac{1}{s} D^T D$$

Then the non-zero eigenvalues of the matrices Σ_S and Σ_D are equal. Moreover, the columns of:

$$\Phi_D = D \cdot \Phi_S$$

where the columns of Φ_S contain the eigenvectors of Σ_S , correspond with the the eigenvectors associated to the non-zero eigenvalues of Σ_D in the sense they have the same direction. Therefore, if the columns of Φ_D are normalized, then Φ_D holds the normalized eigenvectors of Σ_D that has eigenvalues bigger than zero. This not only avoid problems with the memory but also it gives a substantial speed up of the calculations.

3) *Combining color and geometry*: The shape and texture features are concatenated in a matrix (Fig. 1 E). In order to remove correlation between shape and texture and also to make the data representation more compact, a third PCA is performed on the concatenated shape and texture matrix (Fig. 1 F).

4) *Building an individual model*: Once the geometry and texture of the face have been captured, the proposed algorithm builds an individual model for each person in the data set. Each model is built using Fisher linear discriminant analysis. Formally, let X be the data obtained after combining the shape and texture and applying the PCA. Let n_1 be the number of data elements corresponding to the person for whom the model

is being created (class 1) and let n_2 be the number of elements corresponding to the other people (class 2), (Fig. 1 G). Let \bar{x}_1 and \bar{x}_2 be the class mean vectors, \bar{x} be the total mean vector and $x_{i,j}$ be the j th sample in the i th class. Then the between matrix is defined by:

$$\mathbf{B} = n_1(\bar{x}_1 - \bar{x})(\bar{x}_1 - \bar{x})^T + n_2(\bar{x}_2 - \bar{x})(\bar{x}_2 - \bar{x})^T$$

and the within matrix is defined by:

$$\mathbf{W} = \sum_{i=1}^2 \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T$$

The projection that best discriminates the two populations is given by the direction of the eigenvector associated to the maximum eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$ (Fig. 1 H). To ensure that the within matrix \mathbf{W} is not singular, only the f first data variables are taken into account, where f is the number of non-zero eigenvalues of the within matrix \mathbf{W} .

B. classification

In order to obtain a method to classify a given image, the different individual models are firstly standardized so they can be compared. The standardization of model $i = 1, \dots, m$ is based on two assumptions. First, the number of observations for person i is much smaller than the number of the observations of all other people. The second assumption is that the projection of the other people follows a Gaussian distribution. These two assumptions imply that the distribution of all the projected facial images on a particular discriminative individual model can be assumed as a Gaussian distribution with outliers. The standardization of model i is then achieved by transforming the projections into a standard Gaussian distribution, keeping the projections of the person i positive. Formally, let \bar{x}_i be the mean of the projections on model i , σ_i the standard deviation, and let $x_{i,j}$ be the projection of image j in model i . These projections are standardized by:

$$\hat{x}_{i,j} = (x_{i,j} - \bar{x}_i) / \sigma_i$$

If the standardized projection for the images corresponding to person i are negative, then $\hat{x}_{i,j}$ are replaced by $-\hat{x}_{i,j}$ for all projections. This causes the projection of the images corresponding to person i to be positive and far from the mean of the gaussian.

Once that the model i is standardized, the probability of a projected image of belonging to the person i is given by the value of the standard normal cumulative function in the projected value. This fact is used to classify a given image. If it is assumed that the image belongs to a person from the data set, the image is projected by all the models and classified as belonging to the model that gives the largest probability. Moreover, it is also statistically possible to decide if a given person belongs to the data set or it is unknown. This can be achieved by comparing the largest projection obtained in all the models with a probabilistic threshold. E.g, if a 99.9% of probability is required, a given image will only be considered as belonging to the database if the projection in one of the individual models is higher than 3.1 standard deviations.

III. EXPERIMENTAL RESULTS

Two experiments are conducted in order to evaluate the performance of the proposed method. The objective of the first experiment is to evaluate the recognition ability in terms of correct classification rates. This first experiment also aims at ranking the importance of shape and texture. The second experiment aims at analyzing if the proposed method can be incorporated into a biometrical facial recognition scheme. The robustness of the proposed method to the presence of unknowns is considered in this second experiment.

A. Experiment one

The first experiment aims at comparing the performance of the proposed method with respect to the Fisherfaces method in terms of correct classification rates. In order to be consistent with a previously published work [15], unknown people are not taken into account.

To achieve this first goal the AR face database [14] is used. The database is composed of two independent sessions, recorded 14 days apart. At both sessions, each person was recorded 13 times, under various facial poses (all frontal), lighting conditions and occlusions. The size of the images in the database is 768×576 pixels, represented in 24 bits RGB color format.

In this study, a subset of 50 persons (25 male and 25 female) from the database was randomly selected. Seven images per person without occlusions are used from each session. Therefore, the experiment data set is composed of 700 images, with 14 images per person. An example of the selected images for one person is displayed in Fig. 4.



Fig. 4. The AR data set: (Top row) The seven images without occlusions from first session. (Bottom row) The seven images without occlusions from the second session.

All the images were manually annotated with the 22 landmarks previously mentioned.

The data set was divided into two sets. The images of the first session were used to build the individual discriminative models, and images from the second session were subsequently used to test the performance.

The landmarks corresponding to the images in the training set were aligned using full Procrustes analysis. The 44 (x,y)-coordinates were obtained to represent the geometrical configuration of each face. In order to obtain the texture of each face in the training set, the different images were warped with respect to the mean shape. Each of the textures received a histogram equalization in each color band to reduce the differences in global illumination. The textures were converted to gray scale and represented by 41337 pixels. The geometrical

Method	Input features	Correct Classification Rate ¹
MIDM method	Shape	86.4% (95)
MIDM method	Texture	99.6% (3)
MIDM method	Texture and Shape	99.9% (1)
Fishershape	Shape	85.9% (99)
Fisherface	Texture	96.9% (22)
Fishercombined	Texture and Shape	97.1% (20)

TABLE I
CORRECT CLASSIFICATION RATES.

and color representation of each face was combined, reduced and the individual models were built as described in Section II.

The test set was used to evaluate and compare the proposed method with respect to the Fisherface technique. In order to evaluate the importance of the geometrical information, the Fisherface technique was modified replacing the texture data with the shape data and also combining the shape with the texture. These two modified techniques will be referred to as Fishershape and Fishercombined from now on. The Euclidean Nearest-Neighbor algorithm was used as classifier algorithm in the Fisher methods. The proposed method classified the images as the person associated to the model that yields the highest probability.

The test was repeated a second time changing the roles of the training and test sets. So session two was used as training data and session one as test data. The correct classification rates for the different techniques are shown in Table I.

From Table I, it is observed that the proposed method has a slightly better performance than the Fisher methods. Moreover, it is also noticed that using the texture data one obtains a higher accuracy than when the shape is used. This implies that the information contained in the texture is more significant than that included in the shape. However, the information contained in the shape data is not insignificant. The highest correct classification rate in both techniques is attained when both shape and texture are considered.

An interesting property of the proposed algorithm is that it is possible to determine which are the most discriminative features of a given person. In order to illustrate this fact, four models were built using only the texture. The pixels of the faces corresponding to these models which received the 10, 15 and 25% highest weights in the model are displayed (in red) in Fig. 5. It is clear that important discriminating features include eyes, noses, glasses, moles and beards. Notice that the algorithm detects the glasses and the two moles of person 43 as discriminate features.

B. Experiment two

The objective of this second experiment is to test the possibility of incorporating the proposed technique into a biometrical facial recognition scheme. This conveys the identification of people in a data set and also the detection of unknown people. The good performance of the proposed technique in person identification was shown in the previous experiment.

¹Number of misclassified images reported in parentheses.

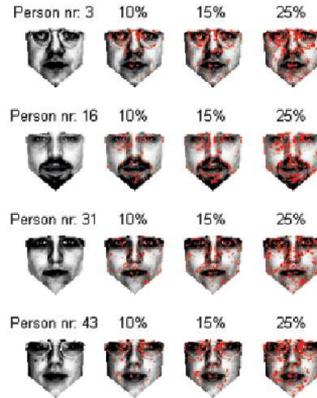


Fig. 5. The 10, 15 and 25% most important pixels (shown in red) for discriminating between the 50 test persons.

Therefore, this second experiment aims at evaluating the performance of the technique in detection of unknown people.

To achieve this goal, the data set used in the previous experiment is selected. In order to evaluate the performance of the technique, a 25-fold crossvalidation was conducted. The seven face images from one male and other seven face images from one female were left out in each iteration. These two people are considered as not belonging to the data set and therefore unknowns. The images of the remaining 48 people were used to train the algorithm.

The False Acceptance Rate (FAR) and False Rejection Rate (FRR) graph, can be observed in Fig. 6. The corresponding Receiver Operating Characteristic curve (ROC) is displayed in Fig. 7.

Both graphs show that the known and unknown populations have a good separability. The best separation happens at the Equal Error Rate (3.1 standard deviations), giving a FAR and FRR of 2%. Moreover, notice that, if the algorithm belongs to a security scheme, the degree of accessibility can be established by increasing or diminishing the standard deviation threshold. E.g., if in the test a false rejection rate of 5.5% is allowed, then a 0% false acceptance rate is obtained. This accommodates biometrical security systems that requires a high level of control access.

A second test is conducted in order to assess the robustness of the proposed method. This test also aims at showing that the method not only discriminates on removable features, such as glasses. To achieve this goal, eight people (four male and four female) are synthetically fitted with four different glasses taken from people belonging to the data set, giving 32 synthetic images.

This second test consists of two steps. First, these eight

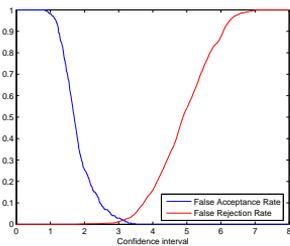


Fig. 6. False Acceptance Rate/False Rejections Rate graph obtained by the 25-fold crossvalidation.

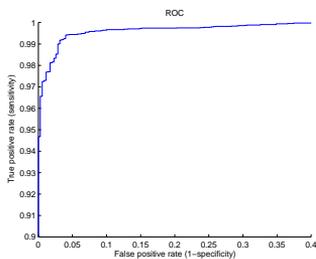


Fig. 7. Receiver Operating Characteristic (ROC) curve obtained by the 25-fold crossvalidation. Notice that only the top left part of the ROC curve is displayed here.

people are not used to built the individual models. The goal is to examine if these eight people who do not belong to the data set are considered as one of the person in the data set. Results show that none of the 32 images is misclassified when a threshold of 3.1 standard deviations is considered (probability of correct classification of 99.9%). This fact can be noticed in Fig. 8 II, where the projections of one of the eight unknown people on the different models are displayed. It is observed that, when the person is considered unknown, his projections onto the individual models belonging to the data set are under the selected threshold. This means that the proposed method does not classify any of the unknown people as belonging to the data set.

In the second step, the eight people (without glasses) are also used to build the individuals models. In this case the goal is to analyze if the method can still recognize people belonging to the data set who has slight changes (same people with glasses). In this second step, the 32 images are also classified correctly by the method. In Fig. 8 III, it is observed that the projections onto the individual model associated with this person clearly surpass the threshold. It is also observed that the projections into the individual models associated to the glasses's owners do not increase significantly. Similar graphs

are obtained for the other seven people. These results show the suitability of the proposed technique in being incorporated into a biometrical security system.

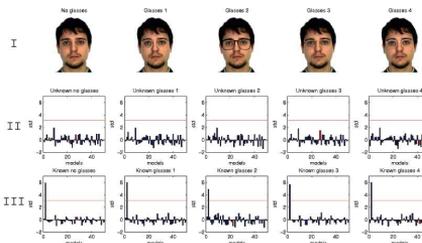


Fig. 8. Impact of changing glasses. (I) Person without glasses and syntetic fitted with 4 glasses form the data set. (II) The corresponding projections in the models as unknown. (III) The corresponding projections in the models as known. Red columns is the model corresponding to the superimposed glasses.

IV. DISCUSSION AND CONCLUSION

In this paper, a novel method to identify people from face images has been proposed. The developed technique aims at being a precise and robust algorithm that can be incorporated into biometrical security systems. The technique has been tested on face images, but it can also be used in other biometrical data, such as speech. Experimental results have proved that the method can attain better classifications rates than an other widely used technique. Moreover, the final one-dimensional projection allows for a simple interpretation of the results. If a given face image is projected onto the different individual models, it is visually possible to determine if this person belongs to one of the models. Moreover, it is also statistically possible to observe the degree of belonging to that model.

Another of the attracting characteristics of the proposed method is its ability to deal with unknowns. The degree of belonging to the data set can be determined statistically. A decision threshold can be determined in relation to a standard Gaussian distribution. This threshold value is used to set the degree of security of the system. The higher this value is set, the smaller the probability of a person being considered as belonging to the data set.

The robustness of the algorithm has been tested using both known and unknown people. The algorithm has been shown to be robust to the inclusion of artifacts such as glasses. On one hand, unknown people using glasses belonging to people from the data set are still classified as unknown. This fact implies that unknown people would not get access to a security system when they use simple removable features belonging to people from the data set. On the other hand, known people using glasses, belonging to other people from the data set, are still recognized as themselves. This means if someone gets glasses, the associated model does not need to

be recalculated. Moreover, this fact suggests that the database should be composed of facial images without glasses. This was also shown by observing that the individual model projections do not change significantly when the glasses were placed.

Another interesting property of the proposed method is its easiness to be maintained and updated. If a large data set is available, it is not needed to recalculate all the existing individual models when a new person has to be registered. Simply, a new individual model for the new person is created. Similarly, if a person has to be removed from the database, it is only needed to remove the corresponding individual model. In conclusion, an accurate, robust and easily adaptable technique to be used for facial recognition has been developed and demonstrated.

REFERENCES

- [1] *Statistical Shape Analysis*. Wiley series in probability and statistics, 1998.
- [2] P.N. Bellhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.
- [3] Hakan Cevikalp, Marian Neamtu, Mitch Wilkes, and Atalay Barkana. Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):4–13, 2005.
- [4] Songcan Chen and Daohong Li. Modified linear discriminant analysis. *Pattern Recognition*, (38):441–443, 2005.
- [5] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, Imaging Science and Biomedical Engineering, University of Manchester, March 2004.
- [6] I. J. Cox, J. Ghosh, and P. N. Yianilos. Feature-based face recognition using mixture-distance. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 209–216, June 1996.
- [7] J. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993.
- [8] J. Daugman. How iris recognition works. *Proceedings of 2002 International Conf. on Image Processing*, 1, 2002.
- [9] G. Finlayson, S. Hordley, G. Schaefer, and G. Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38, 2005.
- [10] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.
- [11] D.G. Kendall. The diffusion of shape. *Advances in Applied Probabilities*, (9):428–430, 1977.
- [12] J. T. Kent. The complex bingham distribution and shape analysis. *Proceedings in current Issues in Statistical Shape Analysis*, pages 167–175.
- [13] S. Z. Li and A. K. Jain. *Handbook of face recognition*. Springer, 2005.
- [14] A.M. Martinez and R. Benavente. The ar face database. Technical Report 24, Computer Vision Center Purdue University, June 1998.
- [15] M. Turk and A.P. Pentland. Face recognition using eigenfaces. *IEEE Conf. Computer Vision and Pattern Recognition*, 1991.
- [16] Ming-Hsuan Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 205–211, 2002.
- [17] A. Yuille, P. Hallinan, and D. Cohen. Feature extraction from faces using deformable templates. *International Journal of computer Vision*, 1992.

APPENDIX C

A face recognition algorithm based on MIDM

A face recognition algorithm based on multiple individual discriminative models

David Delgado Gomez, Jens Fagertun and Bjarne K. Ersbøll

Abstract—In this paper, a novel algorithm for facial recognition is proposed. The technique combines the color texture and geometrical configuration provided by facial images. Landmarks and pixel intensities are used by Principal Component Analysis and Fisher Linear Discriminant Analysis to associate a one-dimensional projection to each person belonging to a reference data set. Each of these projections discriminates the associated person with respect to all other people in the data set. These projections combined with a proposed classification algorithm are able to statistically decide whether a new facial image corresponds to a person in the database. Each projection is also able to visualizing the most discriminative facial features of the person associated to the projection. The performance of the proposed method is tested in two experiments. Results point out that the proposed technique is an accurate and robust tool for facial identification and unknown detection.

I. INTRODUCTION

Regrettable events which happened during the last years (New York, Madrid) have revealed flaws in the existing security systems. The vulnerability of most of the current security and personal identification system is frequently shown. Falsification of identity cards or intrusion into physical and virtual areas by cracking alphanumeric passwords appear frequently in the media. These facts have triggered a real necessity for reliable, user-friendly and widely acceptable control mechanisms for person identification and verification.

Biometrics, which bases the person authentication on the intrinsic aspects of a human being, appears as a viable alternative to more traditional approaches (such as PIN codes or passwords). Among the oldest biometrics techniques, fingerprint recognition can be found. It is known that this technique was used in China around 700 AD to officially certify contracts. Afterwards, in Europe, it was used as person identification in the middle of the 19th century. A more recent biometric technique used for people identification is iris recognition [9]. It has been shown that the chance of finding two randomly formed identical irises is one in 10^{78} (The population of the earth is below 10^{10}) [8]. This technique has started to be used as an alternative to passport in some airports in United Kingdom, Canada and Netherlands. It is also used as employee control access to restricted areas in Canadian airports and in the New York JFK airport. The inconvenience of these techniques is the need for interaction with the individual who wants to be identified or authenticated. This fact has caused that face recognition, a non-intrusive technique, has increasingly received the interest from the scientific community in recent years.

The first developed techniques that aimed at identifying people from facial images were based on geometrical information.

Relative distances between key points, such as mouth corners or eyes, were calculated and used to characterize faces [19]. Therefore, most of the developed techniques during the first stages of facial recognition focused on the automatic detection of individual facial features. However, facial feature detection and measurement techniques developed to date are not reliable enough for geometric feature based recognition, and such geometric properties alone are inadequate for face recognition because rich information contained in the facial texture or appearance is discarded [7], [15]. This fact caused that gradually most of the geometrical approaches were abandoned for color based techniques, which provided better results. These methods aligned the different faces to obtain a correspondence between pixels intensities. A nearest neighbor classifier used these aligned values to classify the different faces. This coarse method was notably enhanced with the appearance of the Eigenfaces technique [17]. Instead of directly comparing the pixel intensities of the different face images, the dimension of these input intensities were first reduced by principal component analysis (PCA). This technique settled the basis of many of the current image based facial recognition schemes. Among these current techniques, Fisherfaces can be found. This technique, widely used and referred [3], [5], combines the Eigenfaces with Fisher linear discriminant analysis (FLDA) to obtain a better separation of the individuals. In Fisherfaces, the dimension of the input intensity vectors is reduced by PCA and then FLDA is applied to obtain a good separation of the different persons.

After Fisherfaces, many related techniques have been proposed. These new techniques aim at providing a projection that attain a good person discrimination and also are robust against differences in illumination or image pose. Kernel Fisherfaces [18], Laplacianfaces [12], or discriminative common vectors [4] can be found among these new approaches. Typically, these techniques have been tested assuming that the image to be classified corresponds to one of the people in the database. In these approaches, the image is usually classified to the person with the smallest Euclidean distance. However, some inconveniences appear when the person to be analyzed may not belong to the data set. In this case, a criterium to decide if the person belongs to the data set has to be chosen. E.g. only people with an Euclidean distance less than a given threshold are considered as belonging to the data set. However, this threshold does not have necessarily to be the same for all the classes (different persons) and different thresholds would need to be found. The estimation of these thresholds is not straightforward and additional data might be needed.

In this work, a new technique that addresses these different

limitations is proposed. The proposed technique introduces two novelties. First, not only the texture intensities are taken into account but also the geometrical information. Second, the data are projected into n one-dimensional spaces instead of a $(n - 1)$ -dimensional space, where n is the number of people in the data set.

Each of these individual models aims at characterizing a given person uniquely. This means that every person in the data set is represented by one model. These multi one-dimensional models allow to statistically interpret the "degree of membership" of a person to the data set and to detect unknowns. Furthermore, these two facts have several advantages in terms of interpretability, characterization, accuracy, and easiness to update the model.

II. ALGORITHM DESCRIPTION

The proposed algorithm is made up of two steps. In the first step, an individual model is built for each person in the database using the color and geometrical information provided by the available images. Each model characterizes a given person and discriminates it from the other people in the database. The second step carries out the identification. A classifier, related with the standard Gaussian distribution, decides whether a face image belongs to one person in the database or not. In this section, the two parts of the algorithm are described in detail. A diagram of the algorithm is displayed in Fig. 1. This diagram will be referred during the description of the algorithm to obtain an easier understanding.

A. Creating the individual models

1) *Obtaining the geometry of the face:* The geometrical characterization of a given face is obtained by means of the theory of statistical shape analysis [2]. In this theory, objects (faces) are represented by shapes. According to Kendall [13], a shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object. In order to describe a shape, a set of landmarks or points of correspondence that matches between and within populations are placed on each face. As an example, Fig. 2A displays a set of 22 landmarks. These landmarks indicate the position of the eyebrows, eyes, nose, mouth, jaw and size of a given face.

To obtain a shape representation according to the definition, the obtained landmarks are aligned in order to remove the location, rotational and scaling effects. To achieve this goal, the 2D-full Procrustes analysis is carried out. Briefly, let:

$$\mathbf{X} = \{\mathbf{x}_i\} = \{x_i + i \cdot y_i\}, \quad i = 1, \dots, n$$

be a set of n landmarks expressed in complex notation. In order to apply full Procrustes analysis, the shapes are initially centered. To center the different shapes, the mean of the shape, $\bar{\mathbf{x}}$, is subtracted from each landmark:

$$\mathbf{w}_i = \mathbf{x}_i - \bar{\mathbf{x}}, \quad i = 1, \dots, n$$

The full Procrustes mean shape [14], $\hat{\mu}$, is found as the eigenvector corresponding to the largest eigenvalue of the

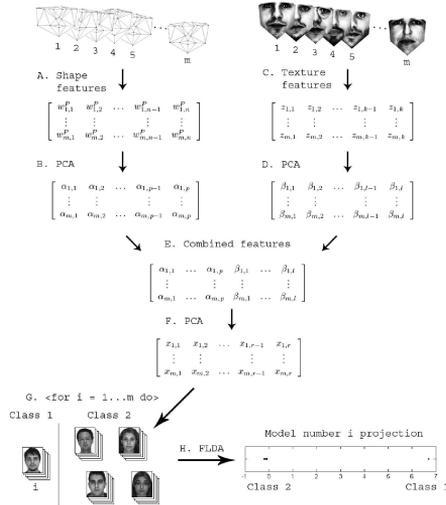


Fig. 1. Algorithm overview. A: Landmarks alignment using full Procrustes analysis. B: PCA on aligned landmarks to remove redundancy. C: Texture normalization using global histogram equalization. D: PCA on normalized texture to remove redundancy. E: Combining shape and texture features. F: PCA on combined features to remove redundancy. G & H: In turn build the individual model using FLDA.

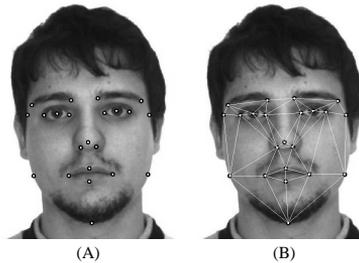


Fig. 2. (A) Set of 22 landmarks placed on a face image. (B) The Delaunay triangulation of the 22 landmarks.

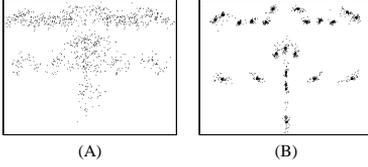


Fig. 3. (A) Superimposition of the sets of 22 landmarks obtained over 49 different face images. (B) Alignment of the landmarks.

complex sum of squares and products matrix

$$\sum_{i=1}^n \mathbf{w}_i \mathbf{w}_i^* / (\mathbf{w}_i^* \mathbf{w}_i)$$

where \mathbf{w}_i^* denotes the transpose of the complex conjugate of \mathbf{w}_i . Using this Procrustes mean shape, the full Procrustes coordinates of $\mathbf{w}_1, \dots, \mathbf{w}_n$ (Fig. 1 A) are obtained by

$$\mathbf{w}_i^P = \mathbf{w}_i^* \hat{\mu} \mathbf{w}_i / (\mathbf{w}_i^* \mathbf{w}_i) \quad i = 1, \dots, n$$

Fig. 3A displays the superimposition of the set of 22 landmarks described in Fig. 2, obtained on 49 different face images. The result obtained after applying the full Procrustes alignment on these landmarks can be observed in Fig. 3B. In order to remove redundancy in the data, a Principal Component Analysis is applied to the aligned landmarks (Fig. 1 B).

2) *Texture formulation*: To form a complete model of the face appearance, the algorithm also captures the texture information provided by the pixels. In order to collect this texture representation, the Delaunay triangulation of every shape is obtained. The Delaunay triangulation connects the aligned landmark set of each image by a mesh of triangles, so no triangle has any of the other points of the set inside its circumcircle. The Delaunay triangulation obtained for each image is warped onto the Delaunay triangulation of the mean shape. The Delaunay triangulation of the 22 landmarks is displayed in Fig. 2B.

Formally, let I be a given image and M the mean shape previously obtained. Let $\mathbf{u}_1 = [x_1, y_1]$, $\mathbf{u}_2 = [x_2, y_2]$ and $\mathbf{u}_3 = [x_3, y_3]$ denote the vertices of a triangle T in I , and let $\mathbf{v}_1, \mathbf{v}_2$ and \mathbf{v}_3 be the associated vertices of the corresponding triangle in M . Given any internal point $\hat{\mathbf{u}} = [x, y]$ in the triangle T , the corresponding point in the associated triangle in the mean shape can be written as $\hat{\mathbf{v}} = \alpha \mathbf{v}_1 + \beta \mathbf{v}_2 + \gamma \mathbf{v}_3$ where:

$$\begin{aligned} \alpha &= 1 - (\beta + \gamma) \\ \beta &= \frac{yx_3 - x_1y - x_3y_1 - y_3x + x_1y_3 + xy_1}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2} \\ \gamma &= \frac{xy_2 - xy_1 - x_1y_2 - x_2y + x_2y_1 + x_1y}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2} \end{aligned}$$

This transformation extracts the texture of a given face image. A histogram equalization is applied to the collected texture to reduce the effects of differences in illumination [11]. This histogram equalization is performed independently in each of the three color channels. Afterwards, the three color

channels are converted into gray scale to obtain a more compact representation (Fig. 1 C).

Similarly to the shape analysis, PCA is conducted in the texture data to reduce dimensionality and data redundancy (Fig. 1 D). However, notice that the large dimension of the texture vectors will produce memory problems because of the huge dimension of the covariance matrix. In order to avoid this difficulty, the Eckart-Young theorem is used [6]. Formally, let \mathbf{D} represents the texture data matrix composed by s n -dimensional texture vectors after the mean of the texture vectors has been subtracted from each one of them ($s \ll n$). Then the $n \times n$ dimensional covariance matrix can be written as:

$$\Sigma_{\mathbf{D}} = \frac{1}{s} \mathbf{D} \mathbf{D}^T$$

Let $\Sigma_{\mathbf{S}}$ be the smaller $s \times s$ dimensional matrix defined by

$$\Sigma_{\mathbf{S}} = \frac{1}{s} \mathbf{D}^T \mathbf{D}$$

Then the non-zero eigenvalues of the matrices $\Sigma_{\mathbf{S}}$ and $\Sigma_{\mathbf{D}}$ are equal. Moreover, the columns of:

$$\Phi_{\mathbf{D}} = \mathbf{D} \cdot \Phi_{\mathbf{S}}$$

where the columns of $\Phi_{\mathbf{S}}$ contain the eigenvectors of $\Sigma_{\mathbf{S}}$, correspond with the eigenvectors associated to the non-zero eigenvalues of $\Sigma_{\mathbf{D}}$ in the sense they have the same direction. Therefore, if the columns of $\Phi_{\mathbf{D}}$ are normalized, then $\Phi_{\mathbf{D}}$ holds the normalized eigenvectors of $\Sigma_{\mathbf{D}}$ that has eigenvalues bigger than zero. This does not only avoid problems with the memory but also gives a substantial speed-up of the calculations.

3) *Combining color and geometry*: Shape and texture features are concatenated in a matrix (Fig. 1 E). In order to remove correlation between shape and texture, and also to make the data representation more compact, a third PCA is performed on the concatenated shape and texture matrix (Fig. 1 F).

4) *Building an individual model*: Once the geometry and texture of the face have been captured, the proposed algorithm builds an individual model for each person in the data set. Each model is built using Fisher linear discriminant analysis. Formally, let \mathbf{X} be the data obtained after combining the shape and texture and applying the PCA. Let n_1 be the number of data elements corresponding to the person for whom the model is being created (class 1) and let n_2 be the number of elements corresponding to the remaining people (class 2), (Fig. 1 G). Let $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ be the class mean vectors, $\bar{\mathbf{x}}$ be the total mean vector and $\mathbf{x}_{i,j}$ be the j th sample in the i th class. Then the between matrix is defined by:

$$\mathbf{B} = n_1 (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}})^T + n_2 (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}})^T$$

and the within matrix is defined by:

$$\mathbf{W} = \sum_{i=1}^2 \sum_{j=1}^{n_i} (\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i) (\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)^T$$

The projection that best discriminates the two populations is given by the direction of the eigenvector associated to the maximum eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$ (Fig. 1 H). To ensure that the within matrix \mathbf{W} is not singular, only the f first data variables are taken into account, where f is the number of non-zero eigenvalues of the within matrix \mathbf{W} .

B. Classification

In order to obtain a method to classify a given image, the different individual models are firstly standardized so they can be compared. The standardization of model $i = 1, \dots, m$ is based on two assumptions. First, the number of observations for person i is much smaller than the number of the observations of all other people. The second assumption is that the projection of the other people follows a Gaussian distribution. These two assumptions imply that the distribution of all the projected facial images on a particular discriminative individual model can be assumed as a Gaussian distribution with outliers. The standardization of model i is then achieved by transforming the projections into a standard Gaussian distribution, keeping the projections of the person i positive. Formally, let \bar{x}_i be the mean of the projections on model i , σ_i the standard deviation, and let $x_{i,j}$ be the projection of image j in model i . These projections are standardized by:

$$\hat{x}_{i,j} = (x_{i,j} - \bar{x}_i) / \sigma_i$$

If the standardized projection for the images corresponding to person i are negative, then $\hat{x}_{i,j}$ are replaced by $-\hat{x}_{i,j}$ for all projections. This causes the projection of the images corresponding to person i to be positive and far from the mean of the gaussian.

Once that the model i is standardized, the probability of a projected image of belonging to the person i is given by the value of the standard normal cumulative function in the projected value. This fact is used to classify a given image. If it is assumed that the image belongs to a person from the data set, the image is projected by all the models and classified as belonging to the model that gives the largest probability. Moreover, it is also statistically possible to decide if a given person belongs to the data set or it is unknown. This can be achieved by comparing the largest projection obtained in all the models with a probabilistic threshold. E.g, if a 99.9% of probability is required, a given image will only be considered as belonging to the database if the projection in one of the individual models is higher than 3.1 standard deviations.

III. EXPERIMENTAL RESULTS

Two experiments are conducted in order to evaluate the performance of the proposed method. The objective of the first experiment is to evaluate the recognition ability in terms of correct classification rates. This first experiment also aims at ranking the importance of shape and texture. The second experiment aims at analyzing whether the proposed method can be incorporated into a biometrical facial recognition scheme. The robustness of the proposed method to the presence of unknowns is considered in this second experiment.

A. Experiment one

The first experiment aims at comparing the performance of the proposed method with respect to the Fisherfaces method in terms of correct classification rates. In order to be consistent with a previously published work [17], unknown people are not taken into account.

To achieve this first goal the AR face database [16] is used. The database is composed of two independent sessions, recorded 14 days apart. At both sessions, each person was recorded 13 times, under various facial poses (all frontal), lighting conditions and occlusions. The size of the images in the database is 768×576 pixels, represented in 24 bits RGB color format.

In this study, a subset of 50 persons (25 male and 25 female) from the database was randomly selected. Seven images per person without occlusions are used from each session. Therefore, the experiment data set is composed of 700 images, with 14 images per person. An example of the selected images for one person is displayed in Fig. 4.



Fig. 4. The AR data set: (Top row) The seven images without occlusions from first session, (Bottom row) The seven images without occlusions from the second session.

All the images were manually annotated with the 22-landmarks previously mentioned.

The data set was divided into two sets. The images of the first session were used to build the multiple individual discriminative models (MIDM), and images from the second session were subsequently used to test the performance.

The landmarks corresponding to the images in the training set were aligned using full Procrustes analysis. The 44 (x,y)-coordinates were obtained to represent the geometrical configuration of each face. In order to obtain the texture of each face in the training set, the different images were warped with respect to the mean shape. Each of the textures received a histogram equalization in each color band to reduce the differences in global illumination. The textures were converted to gray scale and represented by 41337 pixels. The geometrical and color representation of each face was combined, reduced and the individual models were built as described in Section II.

The test set was used to evaluate and compare the proposed method with respect to the Fisherface technique. In order to evaluate the importance of the geometrical information, the Fisherface technique was modified replacing the texture data with the shape data and also combining the shape with the texture. These two modified techniques will be referred to as Fishershape and Fishercombined from now on. The Euclidean Nearest-Neighbor algorithm was used as classifier algorithm

Method	Input features	Correct Classification Rate ¹
MIDM method	Shape	86.4% (95)
MIDM method	Texture	99.6% (3)
MIDM method	Texture and Shape	99.9% (1)
Fishershape	Shape	85.9% (99)
Fisherface	Texture	96.9% (22)
Fishercombined	Texture and Shape	97.1% (20)

TABLE I
CORRECT CLASSIFICATION RATES.

in the Fisher methods. The proposed method classified the images as the person associated to the model that yields the highest probability.

The test was repeated a second time changing the roles of the training and test sets. So session two was used as training data and session one as test data. The correct classification rates for the different techniques are shown in Table I.

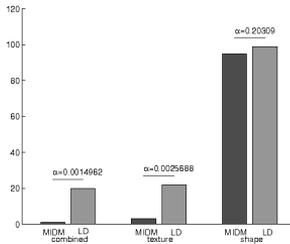


Fig. 5. McNemar Test. Number of misclassified examples by the two classifiers for the combined features, texture features and shape features and the maximum significant level (α) which accept McNemar's null hypothesis.

From Table I, it is observed that the proposed method has a better performance than the Fisher methods. A McNemar's test [1] [10] for each of the three features was carried out in order to show that the observed differences in the performance of both classifiers are significant. Briefly, let n_{10} be the number of examples misclassified by the MIDM classifier but not by the Fisher based method and n_{01} the number of examples misclassified by the Fisher based method but not by the MIDM method for a selected feature. If both classifiers have a similar performance then the statistic

$$\frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}}$$

follows a χ^2 distribution with one degree of freedom. The misclassification errors of both classifiers for each selected feature is displayed in Fig. 5 together with the minimum significant level (α) that will accept the null hypothesis that both classifiers have a similar performance. It is observed that the MIDM classifier has a lesser number of misclassification. Moreover, McNemar test statistically indicates that when the

¹Number of misclassified images reported in parentheses.

texture is considered (as most of the current research projects do), the performance of both classifiers is strongly different.

It is also noticed that using the texture data one obtains a higher accuracy than when the shape is used. This implies that the information contained in the texture is more significant than that included in the shape. However, the information contained in the shape data is not irrelevant. The highest correct classification rate in both techniques is attained when both shape and texture are considered.

An interesting property of the proposed algorithm are that it is possible to determine which are the most discriminative features of a given person. In order to illustrate this fact, four models were built using only the texture. The pixels of the faces corresponding to these models which received the 10, 15 and 25% highest weights in the model are displayed (in red) in Fig. 6. It is clear that important discriminating features include eyes, noses, glasses, moles and beards. Notice that the algorithm detects the glasses and the two moles of person 43 as discriminate features.

B. Experiment two

The objective of this second experiment is to test the possibility of incorporating the proposed technique into a biometric facial recognition scheme. This conveys the identification of people in a data set and also the detection of unknown people. The good performance of the proposed technique in person identification was shown in the previous experiment. Therefore, this second experiment aims at evaluating the performance of the technique in detecting of unknown people.

To achieve this goal, the data set used in the previous experiment is selected. In order to evaluate the performance of the technique, a 25-fold cross validation was conducted. The seven face images from one female were left out in each iteration. These two people are considered as not belonging to the data set and therefore unknowns. The images of the remaining 48 people were used to train the algorithm.

The False Acceptance Rate (FAR) and False Rejection Rate (FRR) graph can be observed in Fig. 7. The corresponding Receiver Operating Characteristic curve (ROC) is displayed in Fig. 8.

Both graphs show that the known and unknown populations have a good separability. The best separation happens at the Equal Error Rate (3.1 standard deviations), giving a FAR and FRR of 2%. Moreover, notice that, if the algorithm belongs to a security scheme, the degree of accessibility can be established by increasing or diminishing the standard deviation threshold. E.g., if in the test a false rejection rate of 5.5% is allowed, then a 0% false acceptance rate is obtained. This accommodates biometrical security systems that requires a high level of control access.

A second test is conducted in order to assess the robustness of the proposed method. This test also aims at showing that the method does not only discriminates on removable features, such as glasses. To achieve this goal, eight people (four male and four female) are synthetically fitted with four different glasses taken from people belonging to the data set, giving 32 synthetic images.

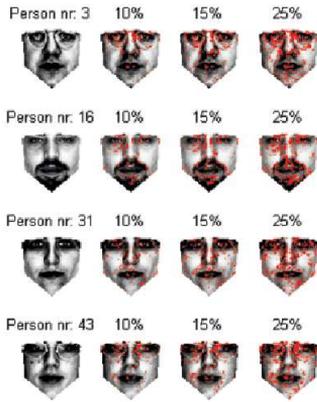


Fig. 6. The 10, 15 and 25% most important pixels (shown in red) for discriminating between the 50 test persons.

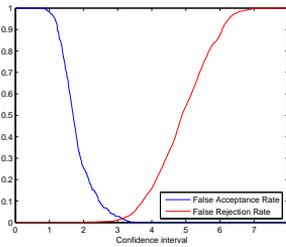


Fig. 7. False Acceptance Rate/False Rejections Rate graph obtained by the 25-fold cross validation.

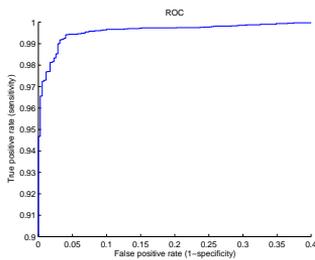


Fig. 8. Receiver Operating Characteristic (ROC) curve obtained by the 25-fold cross validation. Notice that only the top left part of the ROC curve is displayed here.

This second test consists of two steps. First, these eight people are not used to build the individual models. The goal is to examine if these eight people who do not belong to the data set are considered as one of the person in the data set. Results show that none of the 32 images is misclassified when a threshold of 3.1 standard deviations is considered (probability of correct classification of 99.9%). This fact can be noticed in Fig. 9 II, where the projections of one of the eight unknown people on the different models are displayed. It is observed that, when the person is considered unknown, his projections onto the individual models belonging to the data set are under the selected threshold. This means that the proposed method does not classify any of the unknown people as belonging to the data set.

In the second step, the eight people (without glasses) are also used to build the individuals models. In this case the goal is to analyze if the method can still recognize people belonging to the data set who has slight changes (same people with glasses). In this second step, the 32 images are also classified correctly by the method. In Fig. 9 III, it is observed that the projections onto the individual model associated with this person clearly surpass the threshold. It is also observed that the projections onto the individual models associated to the glasses's owners do not increase significantly. Similar graphs are obtained for the other seven people. These results show the suitability of the proposed technique in being incorporated into a biometrical security system.

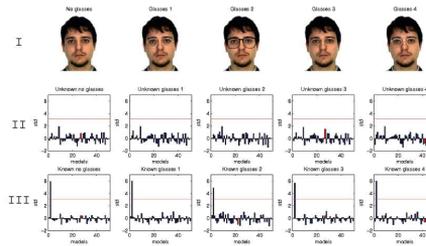


Fig. 9. Impact of changing glasses. (I) Person without glasses and syntetic fitted with 4 glasses from the data set. (II) The corresponding projections in the models as unknown. (III) The corresponding projections in the models as known. Red columns is the model corresponding to the superimposed glasses.

IV. DISCUSSION AND CONCLUSION

In this paper, a novel method to identify people from facial images has been proposed. The developed technique aims at being a precise and robust algorithm that can be incorporated into biometrical security systems. The technique has been tested on face images, but it can also be used in other biometrical data, such as speech. Experimental results have proved that the method can attain better classifications rates than an other widely used techniques. Moreover, the final one-dimensional projection allows for a simple interpretation of the results. If a given face image is projected onto the

different individual models, it is visually possible to determine if this person belongs to one of the models. Moreover, it is also statistically possible to observe the degree of belonging to that model.

Another of the attractive characteristics of the proposed method is its ability to deal with unknowns. The degree of belonging to the data set can statistically be determined. A decision threshold can be determined in relation to a standard Gaussian distribution. This threshold value is used to set the degree of security of the system. The higher this value is set, the smaller the probability of a person being considered as belonging to the data set.

The robustness of the algorithm has been tested using both known and unknown people. The algorithm has been shown to be robust to the inclusion of artifacts such as glasses. On one hand, unknown people using glasses belonging to people from the data set are still classified as unknown. This fact implies that unknown people would not get access to a security system when they use simple removable features belonging to people from the data set. On the other hand, known people using glasses, belonging to other people from the data set, are still recognized as themselves. This means if someone gets glasses, the associated model does not need to be recalculated.

Another interesting property of the proposed method is its easiness to be maintained and updated. If a large data set is available, it is not needed to recalculate all the existing individual models when a new person has to be registered. Simply, a new individual model for the new person is created. Similarly, if a person has to be removed from the database, it is only needed to remove the corresponding individual model. In conclusion, an accurate, robust and easily adaptable technique to be used for facial recognition has been developed and evaluated.

REFERENCES

- [1] *The analysis of contingency tables*. Chapman and Hall, 1977.
- [2] *Statistical Shape Analysis*. Wiley series in probability and statistics, 1998.
- [3] P.N. Bellhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.
- [4] Hakan Cevikalp, Marian Neamtu, Mitch Wilkes, and Atalay Barkana. Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):4–13, 2005.
- [5] Songcan Chen and Daohong Li. Modified linear discriminant analysis. *Pattern Recognition*, (38):441–443, 2005.
- [6] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, Imaging Science and Biomedical Engineering, University of Manchester, March 2004.
- [7] I. J. Cox, J. Ghosh, and P. N. Yianilos. Feature-based face recognition using mixture-distance. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 209–216, June 1996.
- [8] J. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993.
- [9] J. Daugman. How iris recognition works. *Proceedings of 2002 International Conf. on Image Processing*, 1, 2002.
- [10] T. G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, pages 1895–1923, 1998.
- [11] G. Finlayson, S. Hordley, G. Schaefer, and G. Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38, 2005.
- [12] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.
- [13] D.G. Kendall. The diffusion of shape. *Advances in Applied Probabilities*, (9):428–430, 1977.
- [14] J. T. Kent. The complex bingham distribution and shape analysis. *Proceedings in current Issues in Statistical Shape Analysis*, pages 167–175.
- [15] S. Z. Li and A. K. Jain. *Handbook of face recognition*. Springer, 2005.
- [16] A.M. Martinez and R. Benavente. The ar face database. Technical Report 24, Computer Vision Center Purdue University, June 1998.
- [17] M. Turk and A.P. Pentland. Face recognition using eigenfaces. *IEEE Conf. Computer Vision and Pattern Recognition*, 1991.
- [18] Ming-Hsuan Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 205–211, 2002.
- [19] A. Yuille, P. Hallinan, and D. Cohen. Feature extraction from faces using deformable templates. *International Journal of computer Vision*, 1992.

APPENDIX D

FaceRec Quick User Guide

When opening FaceRec one has to select a video driver from which the video input will be received. This is displayed in Figure D.1 and Figure D.2.

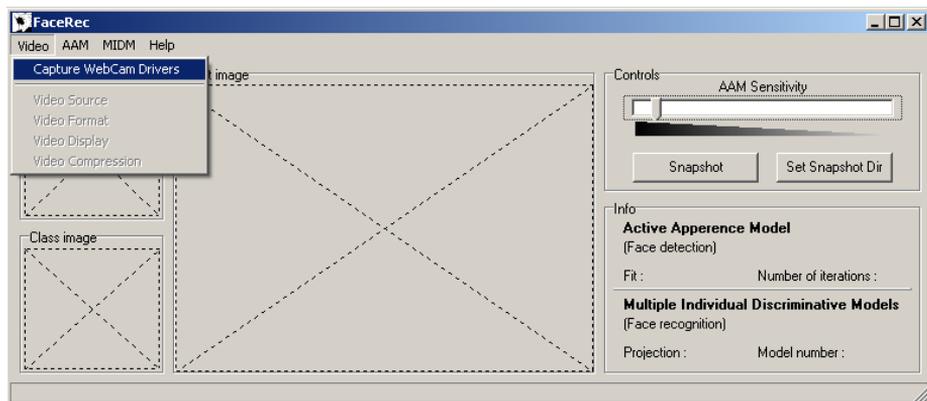


Figure D.1: The main window of the FaceRec application. Go to the menu “Video” -> “Capture WebCam Drivers” to select the video driver.



Figure D.2: The video driver list popup box.

It is possible to adjust the video feed (webcam) by activating the “Video Source”, “Video Format”, “Video Display” and “Video Compression” panel from the “Video” menu, if these are available in the selected video driver (available are denoted by black selectable items and non-available by gray non-selectable items). An example of a Philips PCVC690K webcam “Video Source” panel is displayed in figure D.3.

FaceRec is processing the face detection and face recognition in real time when both an AAM Model File (.amf) and an MIDM Model File (.mmf) are loaded, as displayed in figure D.4.

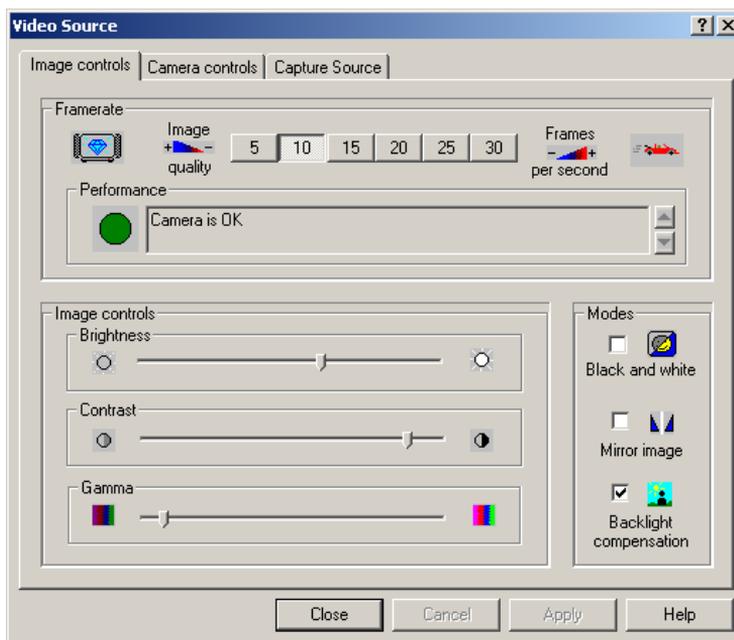


Figure D.3: The specific “Video Source” panel included in the drivers of the Philips PCVC690K webcam.

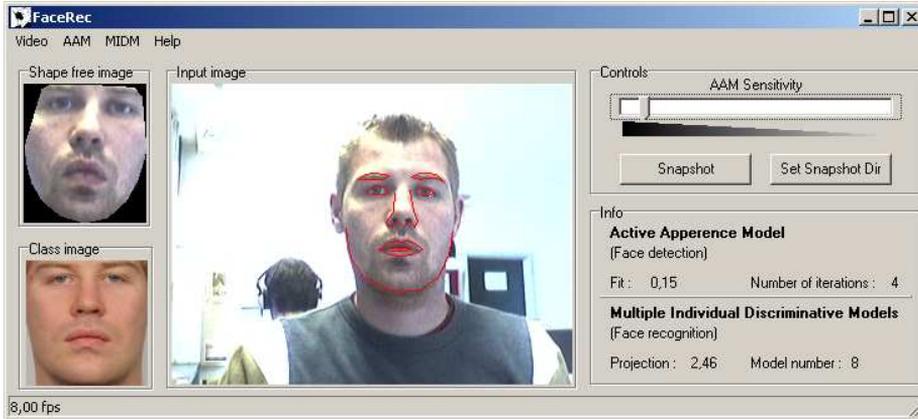


Figure D.4: The FaceRec application capturing a face and recognizing it as model (person) number 8. The FaceRec main window contains the input from the webcam, the shape free image of the captured face, the class image of the recognized person, a control panel and an AAM/MIDM information panel.

In the **Controls** panel it is possible to change the sensitivity of the AAM (face detection). A high sensitivity means that faces are detected with a high accuracy, but it may not detect all faces and vice versa. With the “Snapshot” button it is possible to save snapshots of the video input. If the AAM Model File is loaded the associated shape file (.asf) is also saved to disk. The directory, in which the files are saved, can be changed by the “Set Snapshot Dir”.

The **Info** panel displays information on the fit and the number of iterations used to achieve the fit of the AAM, as well as the largest projection and its model number of this projection in the MIDM. In short, a good recognition is characterized by a low fit and a large projection.

APPENDIX E

AAM-API DLL

DLL index	Function name	Short description
#1	deleteImage	Deletes the AAM image object
#2	deleteModel	Deletes the AAM model object
#3	deleteReferenceFrame	Deletes the AAM reference frame object
#4	deleteShape	Deletes the AAM shape object
#5	evalShape	Manipulates the AAM shape to find a the best fit using a AAM model object and a AAM image object
#6	evalShapeFineTuning	Manipulates the AAM shape to find a the best fit using a AAM model object and a AAM image object
#7	getNCombinedPCA	Returns the number of combined pca parameters of a AAM model object
#8	getNTextureSamples	Returns the number of texture samples in the shape free image of a AAM model object
#9	getShape	Returns the shape coordinates
#10	getShapeBox	Returns the coordinates for the box the AAM shape object resides in
#11	getShapeFreeImage	Returns the shape free image of a AAM shape object, AAM image object and AAM model object
#12	initializeShape	Initialize a AAM shape object

Table E.1: Function list over the `aam_api.dll`.

DLL index	Function name	Short description
#13	modelImage	Models a face (object) using the combined PCA parameters onto an AAM image object
#14	newImage	Constructs a new AAM image object
#15	newModel	Constructs a new AAM model object
#16	newReferenceFrame	Constructs a new AAM reference frame object
#17	newShape	Constructs a new AAM shape object
#18	readImage	Reads an image from a char array into an AAM image object
#19	readModel	Reads an AAM model object from file
#20	shapeCOG	Returns the center of gravity of an AAM shape object
#21	translateShape	Translates an AAM shape object
#22	writeASF	Writes an AAM shape object to file
#23	writeImage	Writes an image from an AAM image object into a char array

Table E.2: Function list over the `aam_api.dll` continued.

APPENDIX F

CD-ROM Contents

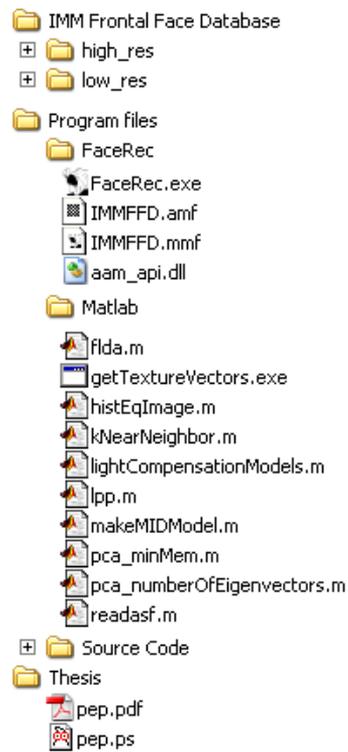


Figure F.1: The contents of the enclosed CD-ROM.

Bibliography

- [1] Jörgen Ahlberg. *Model-based Coding: Extraction, Coding, and Evaluation of Face Model Parameters*. PhD thesis, Department of Electrical Engineering, Linköping University, 2002.
- [2] M. Aizerman, E. Braverman, and L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.
- [3] G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural Computation*, pages 2385–2404, 2000.
- [4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.
- [5] M. Bichsel and A. Pentland. Human face recognition and the face image set's topology. *Computer Vision, Graphics, and Image Processing*, 59(2):254–261, 1994.
- [6] I. Biederman and P. Kalocsai. Neurocomputational bases of object and face recognition. *Philosophical Transactions: Biological Sciences*, 352(1358):1203–1219, 1997.
- [7] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana. Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):4–13, 2005.
- [8] D. Chandler. Visual perception. Internet site (Link : <http://www.aber.ac.uk/media/Modules/MC10220/visindex.html>).

- [9] S. Chen and D. Li. Modified linear discriminant analysis. *Pattern Recognition*, pages 441–443, 2005.
- [10] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, Imaging Science and Biomedical Engineering. University of Manchester, March 2004.
- [11] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.
- [12] I. J. Cox, J. Ghosn, and P. N. Yianilos. Feature-based face recognition using mixture-distance. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 209–216, June 1996.
- [13] I. Craw, D. Tock, and A. Bennett. Finding face features. *Proc. Second European Conf. Computer Vision*, pages 92–96, 1992.
- [14] Y. Dai and Y. Nakano. Face-texture model based on SGLD and its application in face detection in a color scene. *Pattern Recognition*, 29(6):1007–1017, 1996.
- [15] C. Darwin. *The Expression of the Emotions in Man and Animals*. London: John Murray, 1872.
- [16] B. V. Dasarathy. *Nearest Neighbor (NN) Norms : Nn Pattern Classification Techniques*. IEEE Computer Society Press, 1991.
- [17] J. Daugman. How iris recognition works. *Proceedings of 2002 International Conf. on Image Processing*, 1, 2002.
- [18] I. Dryden and K. V. Mardia. *Statistical Shape Analysis*. Wiley series in probability and statistics, 1998.
- [19] B. K. Ersbøll and K. Conradsen. *An introduction to statistics*, volume 2. DTU-tryk, 2003.
- [20] J. Fagertun, D. D. Gomez, B. K. Ersbøll, and R. Larsen. A face recognition algorithm based on multiple individual discriminative models. *Proceedings of the 14th Danish Conference on Pattern Recognition and Image Analysis*, 2005.
- [21] J. Fagertun and M. B. Stegmann. The IMM frontal face database. Technical Report 11, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, aug 2005.

- [22] G. Finlayson, S. Hordley, G. Schaefer, and G. Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38, 2005.
- [23] M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22(1):67–92, 1973.
- [24] F. Galton. Personal identification and description. *Nature*, pages 173–188, 1888.
- [25] I. Gauthier and M. Behrmann. Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience*, 11(4), 1999.
- [26] A. J. Goldstein and L. D. Harmon. Identification of human faces. *Proceedings of the IEEE*, pages 748–760, 1971.
- [27] D. D. Gomez, J. Fagertun, and B. K. Ersbøll. A face recognition algorithm based on multiple individual discriminative models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005. To appear, Submitted in 2005, ID TPAMI-0474-0905.
- [28] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society*, 53(2):285–339, 1991.
- [29] X. He and P. Niyogi. Locality preserving projections. *Proc. Conf. Advances in Neural Information Processing Systems*, 2003.
- [30] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.
- [31] R. Heindl. *System und Praxis der Daktyloskopie und der sonstigen technischen Methoden der Kriminalpolizei*. De Gruyter, Berlin, 1922.
- [32] R. Hietmeyer. Biometric identification promises fast and secure processing of airline passengers. *The International Civil Aviation Organization Journal*, 55(9):10–11, 2000.
- [33] D. G. Kendall. The diffusion of shape. *Advances in Applied Probabilities*, pages 428–430, 1977.
- [34] R. Kjeldsen and J. Kender. Finding skin in color images. *Proc. Second International Conf. Automatic Face and Gesture Recognition*, pages 312–317, 1996.
- [35] A. Lanitis, C. J. Taylor, and T. F. Cootes. An automatic face identification system using flexible appearance models. *Image and Vision Computing*, 13(5):393–401, 1995.

- [36] T. K. Leung, M. C. Burl, and P. Perona. Finding faces in cluttered scenes using random labeled graph matching. *Proc. Fifth IEEE International Conf. Computer Vision*, pages 637–644, 1995.
- [37] S. Z. Li and A. K. Jain. *Handbook of face recognition*. Springer, 2005.
- [38] Q. Liu, R. Huang, H. Lu, and S. Ma. Face recognition using kernel-based fisher discriminant analysis. *Proceedings Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pages 197–201, 2002.
- [39] J. Luetttin and G. Maître. Evaluation protocol for the extended M2VTS database (lausanne protocol). *IDIAP Communication 98-05*, 1998.
- [40] A. M. Martinez. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.
- [41] A. M. Martinez and R. Benavente. The AR face database. Technical Report 24, Computer Vision Center (CVC), June 1998.
- [42] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Czyz, L. Vandendorpe, S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov, R. Paredes, B. Kepenekci, F. B. Tek, G. B. Akar, F. Deravi, and N. Mavity. Face verification competition on the XM2VTS database. *Audio- and Video-Based Biometric Person Authentication. 4th International Conference, AVBPA 2003*, 2003.
- [43] K. Messer, J. Matas, J. Kittler, J. Lüttin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Audio- and Video-based Biometric Person Authentication, AVBPA '99*, pages 72–77, 1999. Washington, D.C., March 1999. 16 IDIAP-RR 99-02.
- [44] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers. Fisher discriminant analysis with kernels. *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pages 41–48, 1999.
- [45] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone. Face recognition vendor test 2002. Technical report, National Institute of Standards and Technology, March 2003.
- [46] F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. Springer, 1985.
- [47] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

- [48] H. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection, 1998.
- [49] K. Skoglund. Three-dimensional face modelling and analysis. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2003. Supervised by Assoc. Prof. Rasmus Larsen.
- [50] M. B. Stegmann. Active appearance models theory extensions and cases. Master's thesis, Technical University of Denmark, 2000.
- [51] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME - a flexible appearance modelling environment. *IEEE Transactions on Medical Imaging*, 22(10):1319–1331, 2003.
- [52] F. Sukno, S. Ordas, C. Butakoff, S. Cruz, and A. Frangi. Active shape models with invariant optimal features IOF-ASMs. In *Proceedings of Audio- and Video-Based Biometric Person Authentication*, pages 365–375, New York, USA, 2005. Springer.
- [53] M. Turk. A random walk through eigenspace. *IEICE Trans. Inf. & Syst.*, E84-D(12):1586–1595, December 2001.
- [54] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, 1991.
- [55] M. Turk and A. P. Pentland. Face recognition using eigenfaces. *IEEE Conf. Computer Vision and Pattern Recognition*, 1991.
- [56] X. Xie and K. M. Lam. Face recognition under varying illumination based on a 2dface shape model. *Pattern Recognition*, 38:211–230, 2005.
- [57] G. Yang and T. S. Huang. Human face detection in complex background. *Pattern Recognition*, 27(1):53–63, 1994.
- [58] J. Yang and A. Waibel. A real-time face tracker. *Proc. Third Workshop Applications of Computer Vision*, pages 142–147, 1996.
- [59] M. H. Yang. Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using kernel methods. *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 205–211, 2002.
- [60] M. H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, January 2002.
- [61] A. Yuille, P. Hallinan, and D. Cohen. Feature extraction from faces using deformable templates. *International Journal of computer Vision*, 1992.