UniS

Computation and Memory Efficient Face Recognition Using Binarized Eigenphases and Component-Based Linear Discriminant Analysis for Wide Range Applications

by

Naser Zaeri

Submitted for the degree of Doctor of Philosophy

Centre for Vision, Speech and Signal Processing
School of Electronics and Physical Sciences
University of Surrey
United Kingdom

June 2007

© Naser Zaeri 2007
Summary

Face recognition finds many important applications in many life sectors and in particular in commercial and law enforcement. This thesis presents two novel methods which make face recognition more practical. In the first method, we propose an attractive solution for efficient face recognition systems that utilize low memory devices. The new technique applies the principal component analysis to the binarized phase spectrum of the Fourier transform of the covariance matrix constructed from the MPEG-7 Fourier Feature Descriptor vectors of the face images. Most of the algorithms proposed for face recognition are computationally exhaustive and hence they can not be used on devices constrained with limited memory; hence our method may play an important role in this area.

The second method presented in this thesis proposes a new approach for efficient face representation and recognition by finding the best location component-based linear discriminant analysis. In this regard, the face image is decomposed into a number of components of certain size. Then the proposed scheme finds the best representation of the face image in most efficient way, taking into consideration both the recognition rate and the processing time. Note that the effect of the variation in a face image, when it is taken as a whole, is reduced when it is divided into components. As a result the performance of the system is enhanced. This method should find applications in systems requiring very high recognition and verification rates. Further, we demonstrate a solution to the problem of face occlusion using this method. The experimental results show that both proposed methods enhance the performance of the face recognition system and achieve a substantial saving in the computation time when compared to other known methods. It will be shown that the two proposed methods are very attractive for a wide range of applications for face recognition.
Acknowledgements

I would like to express my grateful appreciation to my supervisors, Professor Josef Kittler and Dr. Farzin Mokhtarian. During my study, I have received a fruitful supervision from them. The advice, instructions, and instructive comments of both were of great benefits.

Moreover, I would like to thank Professor Abdallah Cherri for his guidance, advice, and valuable discussions.

Finally, my family deserve great gratitude and thanks for their support and continuous encouragement.
Table of Contents

Summary ...................................................................................................................................... ii
Acknowledgements .................................................................................................................. iii
Table of Contents ................................................................................................................... iv
List of Figures........................................................................................................................ vi
List of Tables ........................................................................................................................ viii

Chapter 1  Introduction ................................................................................................. 1
  1.1 Motivation ............................................................................................................. 1
  1.2 Contribution ....................................................................................................... 4
  1.3 Summary of the Thesis .................................................................................... 6

Chapter 2  Literature Survey and Overview of Past Work ........................................ 7
  2.1 Holistic Matching Methods ........................................................................ 7
    2.1.1 Principal Component Analysis ................................................................ 7
    2.1.2 Incremental Principal Component Analysis (IPCA) ................................ 8
    2.1.3 Linear Discriminant Analysis ............................................................. 9
    2.1.4 Independent Component Analysis ..................................................... 11
    2.1.5 Laplacian Faces ................................................................................ 11
    2.1.6 Super-Resolution ............................................................................. 12
    2.1.7 Probabilistic Similarity Measure ....................................................... 13
    2.1.8 Neural Networks ............................................................................. 14
    2.1.9 Support Vector Machine (SVM) ....................................................... 15
  2.2 Feature-Based Matching Methods ................................................................ 16
    2.2.1 Geometrical Features .......................................................................... 16
    2.2.2 Template Matching ............................................................................ 17
    2.2.3 Gabor Wavelets ................................................................................ 18
    2.2.4 Curve-Based Representation ............................................................. 22
    2.2.5 Line Edge Map (LEM) .................................................................... 23
  2.3 Component-based Approaches ..................................................................... 23
  2.4 Other Methods .................................................................................................. 25
    2.4.1 Correlation Filters ............................................................................. 25
    2.4.2 Radial Basis Function (RBF) .......................................................... 26
    2.4.3 Discrete Wavelet Transform ............................................................ 26
    2.4.4 Facial Asymmetry ........................................................................... 27
    2.4.5 Infra-Red Imagery ......................................................................... 27
  2.5 Face Recognition for Limited Memory Applications .................................... 30
  2.6 Conclusion ........................................................................................................ 32

Chapter 3  Pattern Recognition Methods .................................................................. 34
  3.1 Principal Component Analysis ..................................................................... 34
    3.1.1 Mathematical Background .............................................................. 34
    3.1.2 PCA for Face Recognition .............................................................. 37
  3.2 Linear Discriminant Analysis ....................................................................... 40
    3.2.1 Mathematical Background .............................................................. 40
    3.2.2 LDA for Face Recognition ............................................................... 44
  3.3 Conclusion ........................................................................................................ 46
Chapter 4  Limited Memory and High Speed Face Recognition Using Binarized Eigenphases .................................................. 47
  4.1 Introduction .................................................................................................. 47
  4.2 MPEG-7 Fourier Feature Descriptor ........................................................... 49
  4.3 System Description ...................................................................................... 51
  4.4 Experimental Results ................................................................................... 57
  4.5 Conclusions .................................................................................................. 67

Chapter 5  Component-Based Linear Discriminant Analysis ........ 68
  5.1 Introduction .................................................................................................. 68
  5.2 Linear Discriminant Analysis ...................................................................... 70
  5.3 System Description ...................................................................................... 71
  5.4 Feature Selection .......................................................................................... 75
    5.4.1 Filters ................................................................................................... 75
    5.4.2 Wrappers .............................................................................................. 76
  5.5 Experimental Results and Analysis ............................................................. 79
    5.5.1 Occlusion Problem ............................................................................... 82
    5.5.2 Optimality Results for Different Number of Components of Size 20x20 .................................................................................................. 83
  5.6 Conclusion ................................................................................................. 106

Chapter 6  Discussions and Applications ............................................... 107

Chapter 7  Conclusions and Future Work ............................................ 112
  7.1 Conclusions ................................................................................................ 112
  7.2 Future Work ............................................................................................... 114

Appendix  MPEG-7: Multimedia Content Description Interface ................................................................. 117
  A. Introduction ................................................................................................. 117
  B. MPEG-7 Types of Tools ................................................................................ 118
    i. Systems Architecture ................................................................................. 119
    ii. Description Definition Language ............................................................. 120
    iii. Description Schemes ............................................................................. 121
    iv. Descriptors ............................................................................................ 122

References .................................................................................................... 124
List of Figures

Figure 2-1 Super-resolution applied as a pre-processing block to face recognition [51].................................................................13
Figure 2-2 Super-resolution embedded into eigenface-based face recognition [51].................................................................13
Figure 2-3 The basic neural network algorithm used for face detection [3]..............15
Figure 2-4 Horizontal and vertical edge dominance maps [2]......................................17
Figure 2-5 Different regions used in the template matching strategy [2]......................18
Figure 2-6 Gabor Wavelets: the real part of the Gabor kernels at five scales and eight orientations ................................................19
Figure 2-7 Gabor Wavelets: the magnitude of the Gabor kernels at five different scales ........................................................................19
Figure 2-8 The real part of the Gabor wavelet representation of a sample image ........20
Figure 2-9 The magnitude of the Gabor wavelet representation of a sample image .20
Figure 2-10 Nodes and grids for different poses .................................................................................................21
Figure 2-11 The Face Bunch Graph that serves as a general representation of faces. Each stack of disks represents a jet ...................................................................................21
Figure 2-12 Examples of facial surfaces of a person under different facial expressions [58]................................................................................................................22
Figure 2-13 Examples of facial curves for a surface [58]................................................22
Figure 2-14 Correlation outputs for: authentic input (Left) and impostor input (Right). ........................................................................................................26
Figure 2-15 Generic map of superficial blood vessels on the face. (a) Overview of an arterial network. (b) Overview of a venous network. (c) Arteries and veins together underneath the surface of the facial skin [94]......................................28
Figure 2-16 Vascular network extraction [94]................................................................................................................28
Figure 3-1 Converting the N by N image to an N by 1 vector........................................38
Figure 4-1 The process of Fourier Feature extraction for a single image in the MPEG-7 algorithm; Re $[F]$ and Im $[F]$ denotes the real and imaginary parts of the Fourier Transform of the image, and PCLDA refers to the Linear Discriminant Analysis using Principal Component........................................50
Figure 4-2 Block diagram for the new system .........................................................................................52
Figure 4-3 Original covariance matrix $C_s$.........................................................................................53
Figure 4-4 The magnitude part of the Fourier transform of the covariance matrix $C_f$ ........................................................................................................53
Figure 4-5 The phase of the Fourier transform of the covariance matrix $C_f$: Left in Radians, Right in Degrees .........................................................................................54
Figure 4-6 Result after the binarization step.........................................................................................55
Figure 4-7 Principal components ordered by the corresponding eigenvalues from largest to smallest........................................................................................................56
Figure 4-8 Examples from the xm2vts database .........................................................................................59
Figure 4-9 Examples of the images from the ORL database .........................................................................................60
Figure 4-10 The cumulative match scores of the four methods for the ORL database. .........................................................................................62
Figure 4-11 The cumulative match scores of the four methods for the xm2vts database .........................................................................................62
Figure 4-12 The ROC curves of the four methods for the ORL database. .........................................................................................64
Figure 4-13 The ROC curves of the four methods for the xm2vts database. .........................................................................................65
Figure 5-1 The proposed system for the optimal size, number, and locations of component LDA. ............................................................... 72
Figure 5-2 Different images from 4 different sessions (each session with 2 shots) for a certain individual taken from the xm2vts database. ............ 80
Figure 5-3 Examples of the images from the ORL databases. .................. 80
Figure 5-4 A demonstrative example of different components that correspond to a face image. ......................................................... 81
Figure 5-5 Four components face image. .............................................. 82
Figure 5-6 Original, one-quarter, and one-half occluded image. ............... 83
Figure 5-7 Some combinations of the four components face image of size 20×20. .... 84
Figure 5-8 Some combinations of the five components face image of size 20×20. ... 86
Figure 5-9 Some combinations of the six components face image of size 20×20. .... 87
Figure 5-10 Some combinations of the seven components face image of size 20×20. ................................................................. 88
Figure 5-11 A comparison between the recognition rates for different number of components using the xm2vts database. ...................... 89
Figure 5-12 Comparison of the CMS curves of various methods for the xm2vts database (First shot with 1180 images). ................................. 90
Figure 5-13 Comparison of the CMS curves of various methods for the ORL database ................................................................. 91
Figure 5-14 Comparison of the ROC curves of various methods for the xm2vts database (First shot with 1180 images). ............................... 92
Figure 5-15 Comparison of the ROC curves of various methods for the ORL database ................................................................. 93
Figure 5-16 Comparison of the CMS curves of various methods for the xm2vts database (Second shot with 1180 images). ....................... 94
Figure 5-17 Comparison of the ROC curves of various methods for the xm2vts database (Second shot with 1180 images). ....................... 95
Figure 5-18 Comparison of the CMS curves of various methods for the xm2vts database (2360 images). .............................................. 96
Figure 5-19 Comparison of the ROC curves of various methods for the xm2vts database (2360 images). .............................................. 97
Figure 5-20 Relation between the number of eigenvectors encountered and the performance of the system. ........................................... 98
Figure 5-21 Computation time for the LDA process for a single component as a function of window size. ............................................. 100
Figure 5-22 A closer look at the computation time for the LDA process for a single component as a function of window size starting from 30×30. .... 101
Figure 5-23 Four components face image of size 15×15. ................................................................. 103
Figure 5-24 Different combinations of five components face image each of size 15×15. ................................................................. 104
Figure 5-25 Different combinations of six components face image each of size 15×15. ................................................................. 105
List of Tables

Table 4-1 The recognition rates of four different methods..........................61
Table 4-2 Amount of time consumed by each method to perform the training process. 66
Table 4-3 Amount of time consumed by each method to perform the verification of a single image. 67
Table 5-1 The results of the experiments related to Fig. 5.5 (4 components)........82
Table 5-2 Sample results (from the xm2vts database) of the experiments (components) for images in Fig. 5.7. 85
Table 5-3 The results of the experiments related to Fig. 5.8..............................85
Table 5-4 The results of the experiments related to Fig. 5.9..............................88
Table 5-5 The results of the experiments related to Fig. 5.10.............................88
Table 5-6 Comparison of the time consumed (in minutes) and the success recognition rate for the different number of components (size 20×20) of our method and the component LDA described in [22] using 1180 images (First shot) of the xm2vts database. 89
Table 5-7 The time consumed (in minutes) and the success recognition rate for the different number of components and the component LDA method described in [22] using 2360 images of the xm2vts database. 99
Table 5-8 Comparison of the computational complexity between the different components schemes..............................................................102
Table 5-9 The results of the experiments related to Fig. 5.23, the 4 components 15×15 patches.................................................................103
Table 5-10 The results of the experiments related to Fig. 5.24, the 5 components 15×15 patches.................................................................104
Table 5-11 The results of the experiments related to Fig. 5.25, the 6 components 15×15 patches.................................................................105
Chapter 1
Introduction

Although the human brain can analyse, perform segmentation, infer 3D information, classify, recognise and interpret a huge amount of data captured by the eyes, machines can not simply do this process in an easy manner. Trying to automate this process so that it could be performed by a machine turned out to be a very complicated issue that has motivated researchers from various disciplines to spend tremendous effort to accomplish this goal.

The attempts to recognize and identify face images using machines go back four decades, when computers began to impact on our lives. Researchers found that it is important to attempt to understand the strategies that the biological system employs, as a first step towards eventually translating these strategies into machine-based algorithms. These observations provide useful hints that can be valuable to computer vision systems.

1.1 Motivation

Humans have the ability to recognize faces in a cluttered scene with relative ease whereas machine recognition is still a challenge. Face recognition has numerous commercial and non commercial applications. Face recognition has received extensive attention because of its significant applications in many fields, ranging from biometrics, law enforcement and surveillance to smart cards and access control. Moreover, in video processing and internet applications, face image retrieval is an important factor for identification and verification purposes. Accurate automatic personal identification is now needed in a wide range of civilian applications involving the use of passports, cellular telephones, automatic teller machines, and driver licenses.

In our electronically inter-connected society, reliable and user-friendly recognition and verification system is essential in many sectors of our life. The identification and verification of a person can be done in different ways. If the person possesses a physical object (like keys, cards, etc.) then his/her identity can be verified. Further,
the person's knowledge can be used for this purpose, where identity can be verified by checking if the person knows a correct password or a PIN code (Personal Identification Number). Although these techniques are easy to implement, they have some drawbacks: keys can be lost or stolen; passwords can be forgotten or guessed by unauthorised people. Some people write the PIN code of their ATM bank cards on the card itself. Also, in our daily life, the number of passwords that people have to remember in order to take advantage of all services is getting larger. Many people use the same password for all services and when the password is unveiled, the adversary has access to all passwords.

The other important and vital method for identification and verification is using the person's physiological or behavioural characteristics (known as biometrics). Some of these characteristics are the person's fingerprints, hand geometry, face, iris, and voice. The biometric characteristics cannot be misplaced or forgotten. Although biometrics do not solve all security problems and are not appropriate for all applications, they represent powerful means of identity verification in some applications when correctly implemented. Moreover, when combined with the other verification techniques (keys and passwords) some level of security can be reached.

The term biometrics refers to the methods of identifying people based on some aspect of their biology. Fingerprints are a familiar example. However, fingerprints generally require the active participation of the person to be recognized, while a technique such as face recognition could be used to recognize people without their knowledge or cooperation.

In the biometrics context, identification or recognition is the task in which, given a biometric signal, the identity of the person related to the signal is retrieved from a database of N subjects. Here we try to match the biometric signal under inspection to a subject in the database (or reject the signal if it did not pass a certain threshold for any of the N subjects in the database). Identification in this case is a one-to-N matching process.

Another related issue is the verification or authentication, where a biometric signal and a hypothetical identity are given. The task is to verify or ensure that the signal is related to a given identity. Thus biometric data is to be checked against a claimed
identity and in this case we can consider the verification as a one-to-one matching process.

Among all the biometrics that are being used today, the face is the most acceptable one because it is one of the most common methods that humans use in their visual interaction and perception. Despite the great advances in the field of research, face-based identification still poses many challenges. Several images of someone may be dramatically different because of the changes in viewpoint, colour, and illumination, or simply because the person’s looks are different from day to day due to makeup, facial hair, glasses, etc. Faces are rich in information about individual identity but mood, mental state, and position relationships between face parts such as eyes, nose, mouth, and chin affect the performance of face recognition.

Furthermore, the typical face recognition algorithms that found great application in different fields (banks and airports) may not be executable in devices that are memory-constrained, with processor speed of no more than 100-500 MHz. This type of system imposes a great challenge on the memory requirement of the data to be sent. To the best of our knowledge and as we will see in chapter 2, there are few published works that propose solution for devices of limited memory constraints. Chapter 4 of this thesis will discuss this issue and try to propose solutions to deal with such situations.

In addition, chapter 5 of this thesis presents a novel method of component-based face recognition, and a thorough study and comprehensive experiments and results are given there. Most of the systems designed to date can only successfully recognize faces when images are obtained under constrained conditions. One of the important reasons beyond this is that these systems are global or holistic approaches. In other words, they deal with the face image as a whole. In general, holistic approaches suffer from major problems:

1. holistic approaches are not robust to occlusion.
2. they require preliminary segmentation.
3. holistic features are highly sensitive to translation scaling and rotation.
4. these representations are high dimensional.

On the other hand, in a component-based approach, we can allow a flexible geometrical relation between the components, which will help in compensating for
pose changes and other disadvantages of holistic methods. The ability of component-based methods to not require a perfect view of all facial features offers many benefits. In brief, results in the literature strongly suggest that there is a benefit in a component-based approach, and it has attracted researchers' interest recently. However, there are still many difficulties in this field that need to be overcome. Some of these problems are:

- large processing and computational time requirement
- requiring many pre-processing steps
- large image size needed
- selection of the components is done manually and most of them start with components located around pre-selected points on the face.

In this thesis, we will try to solve some of the aforementioned issues. Further, to the best of our knowledge, there is no detailed study discussing the time consumption versus performance. This thesis investigates this relationship and presents such a study.

1.2 Contribution

Motivated by the interesting applications and challenges of face recognition, some of which were mentioned in section 1.1, in this thesis we will study and explore face recognition where we propose solutions to some problems related to this field using novel methods that can cope with some of its difficulties.

Two tracks of research are presented to cover a broad spectrum of applications of face recognition. The first track is introduced in chapter 4 where we propose a binarized eigenphase technique that could solve the problem of face recognition for devices of limited memory and require fast and acceptable recognition rate. As explained in the previous subsection, our main motivation behind this direction of research is the increased commercial interest in portable devices and the need of advanced real-time face recognition systems. For example, in the surveillance systems where the wireless domain for data transmission is adopted, there is a great challenge posted by the memory requirement of the system and the amount of data to be sent. Further, customers keep demanding more cost effective and low power consuming systems.
The main aim beyond this process is to dramatically reduce the size of the space needed to represent the face features, while keeping the performance rate as high as possible. We claim that we managed to achieve some of these targets, as will be seen by the promising results in the tables and curves shown in chapter 4. Further, our proposed system offers another very important advantage: it can function on limited memory devices.

The second track of research is covered in chapter 5 where we present our study of the component-based Linear Discriminant Analysis method. Part of the literature review that will be presented in chapter 2 will cover the component-based approaches in face recognition. Although there is a considerable amount of work in this field, to the best of our knowledge there is no detailed study relating the number of components required to represent the face image, the size of these components, and their locations.

In chapter 5, we will try to tackle the problem of component-based face recognition in a different and new way. We will try to present a study that relates the three factors (number of components, their sizes, and their locations) on one hand, with the performance of the system on the other. Further, we will present some results showing the relationship between the performance of the system and the time consumed by it.

We propose what can be called a “customer need” application. As will be discussed and seen later, the customers will have a range of alternatives or a range of solutions from which they can choose what is suitable for them and consequently they pick the desirable performance versus the corresponding consumed time. Our proposed method gives very good results for recognition and verification when compared to other well known methods. Furthermore, we will show that the proposed technique provides a solution to the occlusion problem in face recognition.

The proposed techniques in chapter 4 and 5 offer, for customers interested in face recognition systems, a wide range of solutions from which they can choose whatever is suitable for their domain of applications. Some of these applications are discussed in some detail in chapter 6.
1.3 Summary of the Thesis

The organisation of the thesis is as follows. In chapter 2 we present a comprehensive literature review for the up-to-date techniques in face recognition field and the most famous ones. We briefly present some of the research efforts in this area. Moreover, research works that attempted to find solutions for face recognition for limited memory applications are investigated in this chapter. Note that this area of research is neglected and a lot of effort is still needed.

Chapter 3 is devoted to the two important and widely used techniques that found great applications in the face recognition area: principal component analysis (PCA) and linear discriminant analysis (LDA). We use both techniques as the basic classification methods in this thesis, and for this reason we furnish a brief mathematical background related to them.

In chapter 4, we present a new approach for face recognition system that can be used for small and portable devices. The main aim of this process is to dramatically reduce the size of the space needed to represent the face features, while keeping the recognition rate as high as possible. The method provides attractive results and can handle the requirement of limited memory and fast processing.

In chapter 5, we present another new method for a face recognition system: the component-based LDA. In this method the face image is partitioned into components and each component separately undergoes LDA transformation. In this chapter we find the optimal size, number, and location of the components that are needed to represent the face image in an efficient way. The technique is shown to be very efficient for face recognition.

Further comments and discussions on the two proposed methods of chapters 4 and 5 are given in chapter 6, where potential and possible applications are also included. The conclusion of the thesis and some possible future directions of this research are given in chapter 7.

Finally we introduce a general knowledge about the well known Multimedia Content Description Interface: MPEG-7 in the appendix. The MPEG-7 describes multimedia data with various descriptors, with face recognition being one of them. The discussion on MPEG-7 descriptors is presented in chapter 4 where we utilize one of these descriptors for a fast and a limited memory face recognition system.
Chapter 2
Literature Survey and Overview of Past Work

In the last 30 years, extensive research has been conducted by psychophysicists, neuroscientists, engineers, and computer vision scientists on various aspects of face recognition by humans and machines. Research on automatic machine recognition of faces started in the 1970s [1, 31, 33]. Many different techniques have been proposed since then, and significant advances have been made in the design of classifiers for successful face recognition. This chapter provides, in some detail, a review of the developments in the field of face recognition where we survey and explore some important techniques proposed by researchers.

2.1 Holistic Matching Methods

These methods use the whole face region as an input to the recognition system. Some of the most famous methods of this approach are given in the following subsections.

2.1.1 Principal Component Analysis

Eigenface-based Face Recognition is one of the most important techniques proposed in face recognition area. It is a holistic method that uses the whole face region as the raw input to a recognition system. When the eigenvectors of the face space are defined, the face images can be represented as a weighted sum of these eigenvectors [10]. The recognition is based upon the process of matching between these weights. The concept of eigenfaces has been extended to represent other features. For example eigeneyes, eigenmouth, and eigennose have been proposed in the matching process. Moreover, eigenfaces that represent different views have been defined for face recognition. Due to its importance and since it has proven to have a great robustness in the face recognition field, and because it has been studied thoroughly and used in our work, we will be discussing this technique in more detail in Chapter 3.
2.1.2 Incremental Principal Component Analysis (IPCA)

The basic PCA process is computationally costly and requires large memory. Every time a new image is added to the gallery set, new calculations for the eigenvectors are required. Usually, PCA is performed in a batch mode, i.e. all training data has to be ready for calculating the PCA projection matrix during the training stage. The learning stops once the training data have been fully processed. If additional training data needs to be incorporated into the existing PCA projection matrix, the matrix has to be retrained with all training data. As a result, it is hard to scale up the developed systems. To overcome this problem and limitation, an incremental approach was investigated and adopted in the community of machine learning. However, the incremental approach has a major limitation: there is no guarantee of bound on the approximation error, since the principal components are obtained sequentially and the computation of the next principal component depends on the previous ones.

The existing IPCA algorithms can be divided into two categories. The first category computed the principal components without computing the covariance matrix [61, 62, 63]. One method to do this is using the candid covariance – free IPCA proposed by Weng et al. [61]. Their method generates “observations” in a complementary space for the computation of the higher order principal components. However, since the principal components are obtained sequentially, the error will be propagated and accumulated in the process.

The second category of the IPCA algorithms reconstructs the significant principal components from the original training images and a newly added sample [64, 65, 66]. When a new sample is added, the dimension of the subspace is increased by one. The updated eigenspace is obtained by using low dimensional coefficient vectors of the original images. Since the dimension of the eigenspace is small, this approach is computationally efficient. However, this approach also suffers from the problem of unpredicted approximation error since the new samples are added one by one and the least significant principal component is discarded to preserve the dimensionality of the subspace.

As an alternative Zhao et al. [67] developed another IPCA method based on the Singular Value Decomposition (SVD) updating algorithm. They have presented a
detailed error analysis and derived an error bound for the approximation between the incremental algorithm and the batch PCA.

The IPCA takes the number of input images, the dimension of the images, and the number of desired non-Gaussian (as they have assumed that the distribution of the random vectors is not a Gaussian distribution) directions as inputs and returns the image data matrix and the non-Gaussian vectors as outputs. The IPCA works like a linear system that predicts the next state vector from an input vector and a current state vector. The non-Gaussian components will be updated from the previous component values and from a new input image vector. Whereas the traditional PCA algorithm computes eigenvectors and eigenvalues for a sample covariance matrix derived from a well-known given image data matrix by solving an eigenvalue system problem, the IPCA method updates the eigenvectors each time a new image is introduced.

Researchers in [52], merge sequentially two techniques based on Principal Component Analysis and Independent Component Analysis (ICA). However, they have not used the popular PCA. Instead, they have used IPCA. While IPCA returns the estimated eigenvectors as a matrix that represents subspaces of data, the ICA searches for the independent directions where the projections of the input data vectors will maximise the non-Gaussianity.

One important and major difference between the IPCA-ICA algorithm and the PCA-ICA algorithm is the real-time sequential process nature of the former. Thus, the IPCA-ICA does not need a large memory to store the whole data matrix that represents the incoming images, since the next incoming images will be stored over the previous images and vectors.

### 2.1.3 Linear Discriminant Analysis

Etemad et al. [12] proposed the famous LDA-based feature extraction for face recognition. They have explained that PCA provides the features that capture the main directions along which the face images differ the most, but the within class scatter of the feature points are reduced. However, LDA uses the information of each class (individual) and finds the representation in a way that emphasizes the variations between the various classes and de-emphasizes the variations between images of the
same class. The major problem with the LDA method is the singularity of the within class scatter matrix. Belhumeur et al. [21] has proposed a solution for the problem of singularity of the within class scatter matrix by first projecting the image into a lower dimensional space using the PCA technique and then applying the LDA. This will ensure that the within class scatter matrix is non-singular. LDA has been extensively used for face recognition in last few years. LDA works have been proposed to overcome some of the problems associated with its application to face recognition and to ensure better performance [77, 83, 100]. Since LDA plays a major role in our work, we will discuss it in more detail in Chapter 3.

Zheng et al. [100] provided mathematical theory that proves the feasibility of using PCA as a dimensionality reduction process in the LDA. They have discussed the benefits of adding some smaller principal components instead of some larger ones. They have conducted some experiments and intuitively observed that some of the smaller principal components have better balance on maximizing the between-class distance while minimizing the within-class distance, than the selected large principal components. Moreover, although PCA has been employed in face recognition technology for more than 15 years, there is no rigid algorithm for determining which principal components should be used for face recognition. As a solution for this problem, they proposed Generic Algorithm method for principal component selection [100], and consequently they have integrated it with the LDA technique.

Kim et al. [30] proposed a method of nonlinear discriminant analysis involving a set of locally linear transformations called Locally Linear Discriminant Analysis (LLDA). Their idea was stimulated by the assumption that the global nonlinear data structures are locally linear and local structures can be linearly aligned. As they explained, their method is computationally more efficient compared with the conventional nonlinear methods that are based on kernels, because their method involves only linear transformations. However, it was argued by [31] that this method faces some difficulties. It relies on accurate landmark detection. The true transformation is nonlinear and subject dependent, and more importantly the method cannot recover the information lost due to lack of visibility.
2.1.4 Independent Component Analysis

PCA can only separate pair-wise linear dependencies between pixels. Higher order dependencies will still show in the joint distribution of PCA coefficients, and will not be properly separated. In the task of face recognition, much of the important information may be contained in the higher order relationships among the image pixels [42]. ICA is one of the generalizations of the PCA that is sensitive to high-order statistical relationships. A number of algorithms for performing ICA have been proposed [43, 44]. In [42], they have employed an algorithm developed by Bell and Sejnowski [45], the infomax algorithm derived from the principle of optimal information transfer in neurons with sigmoidal transfer functions. PCA can be derived as a special case of ICA when Gaussian source models are used. In this case, the likelihood of the data depends only on the first- and second - order statistics. The assumption of Gaussian sources implicit in PCA makes it inadequate when the true sources are non - Gaussian. It has been argued that face images can be better described using ICA as it can provide a better probabilistic models and it can find a not-necessarily orthogonal basis which may reconstruct the data better than PCA. Moreover, ICA is sensitive to high-order statistics in the data. So it can represent the face space better in a noisy environment.

In [59], a Fisher version of the ICA is regarded as a systematic classification framework where they compared the InfoMax algorithm proposed by Bartlett [42] with the FastICA algorithm proposed by Hyvarinen [60] (that finds a direction such that the projection maximizes the non-Gaussianity). The proposed approach in [59] consists of three stages. The first stage is used to project face pattern from a high dimensional image space into a space with lower dimensionality using the PCA technique. In the second stage, they invoke the ICA algorithm to find statistically independent basis images while the third stage is applied to the linear discriminant analysis method to take advantage of class-specific information.

2.1.5 Laplacian Faces

Recently, a number of researches have shown that face images possibly reside on a nonlinear submanifold [70, 71, 72, 73]. Both PCA and LDA fail to discover these types of structures. Thus nonlinear techniques have been proposed to deal with the nonlinear structure of the manifold. Examples of these techniques are Isomap, Locally
Linear Embeddings, and Laplacian Eigenmap [74, 75, 76]. In addition, kernel eigenfaces and kernel Fisherfaces techniques can discover the nonlinear structure of the face images; although they are computationally expensive [77]. Another approach for face representation and recognition that explicitly considers the manifold structure is presented in [78]. The manifold structure is modeled by a nearest – neighbor graph which preserves the local structure of the image space. A face subspace is obtained by Locality Preserving Projections, where each face image in the image space is mapped to a low dimensional face subspace that is characterized by a set of feature images, called Laplacian faces. The face subspace preserves local structure and seems to have good discriminating power when compared to PCA. Therefore, Locality Preserving Projection is an algorithm for learning a locality preserving subspace that seeks to maintain the inherent geometry of the data and local structure. It is important to mention that the basis functions obtained by the Laplacian face method are non-orthogonal, which makes it difficult to reconstruct the data. Cai et al. [79] proposed an orthogonal Laplacian face algorithm, which is fundamentally based on the Laplacian face method. The orthogonal Laplacian face method shares the same locality preserving character as Laplacian face, but at the same time it requires the basis functions to be orthogonal. It has been stated in [79] that orthogonal Laplacian face have more locality preserving power than the Laplacian face, and consequently it is expected to have more discriminating power.

2.1.6 Super-Resolution

In the past, super-resolution techniques [48, 49, 59] have been proposed that attempt to increase the resolution by combining information from multiple images. These techniques use super-resolution as a pre-processing system to obtain a high resolution image that can later be passed to a face recognition system. Considering that most state-of-the-art face recognition systems use an initial dimensionality reduction method, [47, 51] proposed embedding the super-resolution algorithm into the face recognition system so that the super resolution is not performed in the pixel domain, but is instead performed in a domain of reduced dimensionality. By this the computation complexity is reduced. In a real-time surveillance scenario where the super-resolution algorithm is expected to work on continuous video streams, computational complexity is usually a very critical issue. In other words, the face
observations are first projected to the face space, and then the super-resolution reconstruction is performed in the low-dimensional face subspace instead of the spatial domain. Figs. 2.1 and 2.2 show super-resolution applied as a pre-processing block, and when embedded into eigenface-based face recognition, respectively.

Figure 2-1 Super-resolution applied as a pre-processing block to face recognition [51].

Figure 2-2 Super-resolution embedded into eigenface-based face recognition [51].

2.1.7 Probabilistic Similarity Measure

Probabilistic methods are also used in face recognition. For instance, image matching for visual object recognition and image database retrieval make use of different image similarity measures such as the probabilistic similarity measure [4]. This similarity measure is based on a Bayesian analysis using two mutually exclusive classes of image variation as encountered in a typical face recognition task. The high-dimensional probability density functions, one for each class, are obtained from training data using an eigenspace density estimation technique. They are subsequently used to compute a similarity measure based on the relevant probability, which is used to rank matches in the database. The probabilistic similarity measure is based on the probability that the image intensity differences, denoted by $\Delta = I_1 - I_2$, are characteristic of typical variations in appearance of the same object. For purposes of
face recognition, the two classes of facial image variations are defined as: *intrAPERsonal* variations denoted by $\Omega_i$, (corresponding, for example, to different facial expressions of the same individual) and *extrAPERsonal* variations denoted by $\Omega_E$, (corresponding to variations between different individuals). The similarity measure is then expressed in terms of the probability

$$S(I_1, I_2) = P(\Delta \in \Omega_i) = P(\Omega_i | \Delta)$$

(2.1)

where $P(\Omega_i | \Delta)$ is the *posterior* probability given by Bayes rule, using estimates of the likelihoods $P(\Delta | \Omega_i)$ and $P(\Delta | \Omega_E)$ that are derived from training data using an efficient subspace method for density estimation of high-dimensional data. It is assumed that both classes are Gaussian-distributed.

Given the likelihoods $P(\Delta | \Omega_i)$ and $P(\Delta | \Omega_E)$, the similarity score $S(I_1, I_2)$ between a pair of images is defined, using Bayes rule, as

$$S(I_1, I_2) = P(\Omega_i | \Delta) = \frac{P(\Delta | \Omega_i) P(\Omega_i)}{P(\Delta | \Omega_i) P(\Omega_i) + P(\Delta | \Omega_E) P(\Omega_E)}$$

(2.2)

where the priors $P(\Omega)$ can be set to reflect operating conditions (e.g., number of test images vs. the size of the database) or other sources of a *prior* knowledge regarding the two images being matched. One method for density estimation involves dividing the vector space $\mathbb{R}^N$ into two complementary subspaces using eigenspace decomposition [4]. This method relies on a PCA to form a low-dimensional estimate of the complete likelihood which can be evaluated using only the first $M$ principal components, where $M << N$.

### 2.1.8 Neural Networks

Neural networks found extensive applications in face detection and recognition field. Authors in [3] presented a neural network-based algorithm to detect upright, frontal views of faces in grey-scale images. The algorithm works by applying one or more neural networks directly to portions of the input image and arbitrating their results. Each network is trained to output the presence or absence of a face. The algorithms and training methods are designed to be general, with little customisation for faces.
Unlike face recognition, in which the classes to be discriminated are different faces, the two classes to be discriminated in face detection are "images containing faces" and "images not containing faces." It is easy to get a representative sample of images which contain faces, but much harder to get a representative sample of those which do not. The proposed system operates in two stages: It first applies a set of neural network-based filters to an image and then uses an arbitrator to combine the outputs. The filters examine each location in the image at several scales, looking for locations that might contain a face. The arbitrator then merges detections from individual filters and eliminates overlapping detections. Fig. 2.3 shows basic algorithm for face detection using neural network.

Moreover, the popular back-propagation neural net may be trained to recognize face images directly [11, 3]. However, for a moderate image size (for example, 128×128) the number of the inputs to the network would be very high. In order to reduce the complexity, two back-propagation networks are used where the first network operates in the auto-association mode and extracts the feature for the second network, which operates in the classification mode.

![Figure 2-3](image)

**Figure 2-3** The basic neural network algorithm used for face detection [3].

### 2.1.9 Support Vector Machine (SVM)

The foundations of Support Vector Machines (SVM) have been developed by Vapnik [134] and gained popularity due to many attractive features. The formulation embodies the Structural Risk Minimization principle, which has been shown to be superior [164] to traditional Empirical Risk Minimization principle, employed by conventional neural networks. Structural Risk Minimization minimizes an upper
bound on the Vapnik Chervonenkis (VC) dimension (generalization error), as opposed to Empirical Risk Minimization that minimizes the error on the training data. SVM were developed to solve the classification problem, where it can be used to train polynomial, neural network, or radial basis function classifiers. Training a SVM is equivalent to solving a linearly constrained quadratic programming problem in a number of variables equal to the number of data points [132, 133, 134]. Jonsson et al. [34] implemented SVM to face verification and recognition where their study supported the hypothesis that the SVM is able to extract the relevant discriminatory information from the training data and its performance outperformed the eigenface method.

2.2 Feature-Based Matching Methods

This approach uses the local features of the face image such as the eyes, nose, and mouth for representation. These parts are supposed to be the most important parts of the face that can characterize the face space properly. The geometry and the appearance of these features are fed into the recognition system. The following subsections present some of the important methods of this approach.

2.2.1 Geometrical Features

Authors in [2] investigated and compared the performance of face recognition based on two traditional classes of techniques applied to the recognition of digital images of frontal views of faces under roughly constant illumination. The first class of techniques is based on the computation of a set of geometrical features from the image of a face. The second class of techniques is based on template matching. The overall geometrical configuration of the face is described by a vector of numerical data representing the position and size of the main facial features: eyes and eyebrows, nose, and mouth. The shape of the face outline can supplement this information. The extracted features must be somehow normalized in order to be independent of position, scale, and rotation of the face in the image plane. A very useful technique for the extraction of facial features is the integral projections. Let $I(x,y)$ be the image. The vertical and horizontal integral projection of $I(x,y)$ in the $[x_1,x_2] \times [y_1,y_2]$ rectangle are defined as
respectively. Partitioning the edge map in terms of edge directions performs an edge projection analysis. As it is described in [2], there are two main directions in the constrained face images: horizontal and vertical. A pixel is considered to be in the vertical edge map if the magnitude of the vertical component of the gradient at that pixel is greater than the horizontal one. The gradient is computed using a Gaussian regularization of the image. Only points where the gradient intensity is above an automatically selected threshold were considered. Fig. 2.4 shows the horizontal and vertical edge dominance maps.

Horizontal gradients are useful to detect the left and right boundaries of face and nose, whereas vertical gradients are useful to detect the head top, eyes, nose base, and mouth. After the detection of the facial features is done, the recognition process is performed with a Bayes classifier.

### 2.2.2 Template Matching

Another method implemented in [2] is based on the use of whole image grey-level templates. Each person is represented by a database entry whose fields are a digital image of his/her frontal view and a set of four masks representing eyes, nose, mouth, and face, as shown in Fig. 2.5.
When attempting recognition, the unclassified image is compared, in turn, with all of the database images, returning a vector of matching scores (one per feature) computed through normalized cross correlation. The unknown person is then classified as the one giving the highest cumulative score.

Deformable templates technique is another method for detecting and describing features of faces [35]. Deformable templates are similar to classical correlation templates with some built-in non-rigidity component. One deformable template approach uses parameterised curves and surfaces to model the non-rigid elements of faces and facial subfeatures, such as the eyes, nose, and lips. The parameterised curves and surfaces are fixed elastically to a global template frame to allow for minor variations in position between facial features. The matching process aligns the parameterised curves and surfaces to their corresponding image features, while minimizing deformation “stress” in the template. The feature of interest, an eye for example, is described by a parameterised template, which enables a priori knowledge about the expected shape of the features to guide the detection process. An energy function is defined which links edges, peaks and valleys in the image intensity to corresponding properties of the template. The template then interacts dynamically with the image, by altering its parameter values to minimize the energy function, thereby deforming itself to find the best fit. The final parameter values can be used as descriptors for the feature.

2.2.3 Gabor Wavelets

Introduced to image analysis due to their biological relevance and computational properties, the Gabor wavelet representation captures salient visual properties and
captures the local structure corresponding to spatial frequency, special localization, and orientation selectivity [80, 81, 82]. As a result, the Gabor wavelet representation should be robust against variation in illumination and facial expression changes. The Gabor wavelets (kernels, filters) are all self similar since they are generated from one filter, the mother wavelet, by scaling and rotation via the wave vector. Figs. 2.8 and 2.9 show the Gabor wavelet representation (the real part and the magnitude, respectively) of a sample image. These representation results display scale, locality, and orientation properties corresponding to those displayed by the Gabor wavelets in Figs. 2.6 and 2.7.

Gabor wavelets are generated from one “mother” kernel, by defining different scaling and rotation factors [83, 84, 113, 125]. Then, each face image is convolved with a family of these kernels and the outputs of these convolutions are concatenated to represent a feature vector for a certain class.

![Figure 2-6 Gabor Wavelets: the real part of the Gabor kernels at five scales and eight orientations.](image)

![Figure 2-7 Gabor Wavelets: the magnitude of the Gabor kernels at five different scales.](image)
Wiskott et al. [125] presented a system for face recognition called Elastic Bunch Graph Matching. The faces are represented by labelled graphs, based on a Gabor wavelet transform. A labelled graph $G$ representing a face consists of $N$ nodes connected by $E$ edges. The nodes are located at facial landmarks called fiducial points, e.g., the pupils, the corners of the mouth, the tip of the nose, the top and bottoms of the ears, etc. The nodes are labelled with jets $J_n$, where these jets are based on a wavelet transform defined as a convolution of the image with a family of Gabor kernels

$$
\Psi(k, x) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2 x^2}{2\sigma^2}\right) \left[\exp(ik \cdot x) - \exp\left(-\frac{\sigma^2}{2}\right)\right]
$$

(2.5)
where \( a = 2\pi \) and \( k \) is the wave vector. Graphs for different head pose differ in geometry and local features (jets). A representative set of \( M \) individual model graphs are combined into a stack-like structure, called a face bunch graph. A set of jets referring to one fiducial point is called a bunch. Fig. 2.10 shows the nodes and grids for different poses of a face image. Fig. 2.11 shows the face bunch graph that serves as a general representation of faces. Note that each stack of disks represents a jet.

C. Liu [83] introduced a Gabor – Fisher Classifier method for face recognition where he applies the enhanced Fisher linear discriminant model to augmented Gabor feature vector derived from the Gabor wavelet representation of face images. To encompass all the features produced by the different Gabor kernels, one concatenates the resulting Gabor wavelet features to derive an augmented Gabor feature vector. The dimensionality of the Gabor vector space is then reduced under the eigenvalue selectivity constraint of the enhanced Fisher linear discriminant method to derive a low-dimensional feature representation with enhanced discrimination power. The proposed algorithm improves the generalization capability of Fisher linear
discriminant method by decomposing the Fisher linear discriminant procedures into a simulation diagonalization of the two within- and between-class scatter matrices. The simultaneous diagonalization is step-wisely equivalent to two operations: whitening the within-class scatter matrix and applying PCA on the between-class scatter matrix using the transformed data [123].

2.2.4 Curve-Based Representation

Moreover, a number of approaches have emerged for comparing shapes of facial surfaces using 2D and 3D shapes [56, 57, 58], where a set of features (nose, nose bridge, eyes, lips, etc) and their locations in a face are detected, and their relative placements are used to characterize the facial shape. In [58], Samir et al. derived curve-based representation of facial surfaces and imposed metrics that compare these representations. They represented a facial surface as a union of closed, planer curves, called facial curves, and compared facial surfaces implicitly by comparing the corresponding facial curves. Curves on surfaces are defined using level sets of real-valued functions on these surfaces. The facial surface is captures by 3D scans and is represented by a polygonated mesh with a collection of connected edges and vertices, while the facial curve is represented by depth function. Fig. 2.12 shows examples of facial surfaces of a person under different facial expressions, whereas Fig. 2.13 shows examples of facial curves for a certain surface.

![Figure 2-12 Examples of facial surfaces of a person under different facial expressions [58].](image1)

![Figure 2-13 Examples of facial curves for a surface [58].](image2)
2.2.5 Line Edge Map (LEM)

LEM is another new method proposed for face recognition based on encoding faces into binary edge maps using Sobel edge detection algorithm. The structural information is integrated with spatial information of a face image by grouping pixels of face edge map to line segments. Then, a polygonal line fitting process is applied to generate the LEM of a face [126]. To match between faces, a measure that is based on Line Segment Hausdorff Distance is used.

2.3 Component-based Approaches

Face recognition using component-based approach has attracted some focus in the last few years [22, 137, 138, 139, 140, 141, 142, 146, 147, 148, 149, 150, 151, 153, 155, 157, 158, 159]. At the first stage the approach finds local features of the face image using a set of characteristics of the face. Then the results of the local features are combined at the second stage to decide whether the input face image belongs to a given class. The early work in this direction was already discussed in this chapter. Brunelli et al. [2] identified faces by independently matching templates of three facial regions: eyes, nose, and mouth. Moghaddam et al. [4] applied PCA to local facial components. Also, the algorithm described by Wiskott et al. [125] uses the local features for identification, where they have applied Gabor filters to extract the face features. One of the main objectives of the component-based approach is to find the best set of components, including their locations, their number, and their size that can classify and identify the face image.

Kim et al. [146] proposed a combined subspace method using both global and local features where the combined subspace is evaluated in view of the Bayes error. They have used several pre-processing steps, besides the resolution of the face images used was 100×120. As a result, their method required more computational time. Ullman et al. [138] used both whole face images at intermediate resolution and local regions at high resolution for face verification, where components of various sizes were cropped at random locations of the images. Weber et al. [147] use localized image patches and explicitly compute their joint spatial probability distribution. Viola and Jones [148] select rectangular features with AdaBoost trained classifier. Chen et al. [149] also use this boosting approach for components learned by local non-
negative matrix factorization. Amit and Geman [150] employ small, localized and oriented edges and combine them with decision trees.

Dorko et al. [139] proposed a method for selecting discriminative scale-invariant object parts and characterize each of them by a scale, rotation, and illumination invariant descriptor. Their method, which was implemented for car recognition, used two types of classifiers. Support Vector Machines and classification based on Gaussian mixture model. The components were selected based on mutual information and likelihood ratio. They applied a generic interest operator to automatically determine an initial set of components located in the vicinity of the detected points of interest. However, forcing the locations of the components to coincide with points already detected may lead to the loss of important features. Similarly, Lowe [142] applied a strategy to automatically determine an initial set of components by applying a generic interest operator to the training images to pick components located in the surroundings of the detected important points. Again, their strategy restricts the choice of possible components that might be important for classification.

Heisele et al. [140] introduced an algorithm that learns rectangular facial components for a face detection system. The algorithm starts with small initial components located around pre-selected points on the face, and each component is grown iteratively. In [151], authors extended this method to the multi-class problem of face recognition. This method required extraction of components from a large number of training images, and consequently it was computationally costly. Again, in [152], Heisele et al. used a set of seven textured 3-D head models with known point-wise 3-D correspondences for the extraction process. They had to locate the maxima of the responses of the component classifiers and compute the output of the geometrical classifier. Further, Heisele et al. in [158] restricted the location of the components to be within a pre-defined search region and they add the position of the detected pre-defined components as an additional input to the classifier. Also, they have used linear and second-degree polynomial SVMs classifiers in their system, which make their system computationally exhaustive. As a result, their method was again computationally expensive.

Ivanov et al. [153], trained individual recognition classifiers on each extracted component and merge their outputs using different combination strategies. However, the classifiers that were used are Support Vector Machines with polynomial kernel of
degree 2. A probabilistic approach using part-selected matching has been proposed in [154] for expression invariant and occlusion tolerant recognition of frontal faces of size 120×170. Mohan [137] manually selected five parts of the human body for person identification. Kim et al. [22] apply component-based LDA onto facial components and offer a good representation for MPEG-7 face description. However, in their study they do not explain how the facial components are chosen. Further, no criteria are mentioned on how to choose the size of component that should be used, or how many components are needed, or the proper location of these components. Moreover, the time consumed by their method is very large.

Short et al. [155] presented experiments in component-based face verification, where they have compared the performance of different photometric normalization techniques of varying complexity on sub-images containing salient features of the face. Mu et al. [156] adapted the process by which the graph localizes itself on the face image allowing the graph to deform in a hierarchical way. Also, authors in [157] presented a method of describing images in terms of local features derived from a PCA representation.

2.4 Other Methods

2.4.1 Correlation Filters

The idea of spatial frequency domain (or correlation filters as it is usually called) object recognition in not new. Many researchers investigated the idea of spatial frequency domain methods for automatic target recognition, where they tried to locate and classify various targets in a scene [102, 104, 105, 106, 107]. It has proved its efficiency in dealing with image variability, and there is much effort to implement this idea to deal with face image variability. However, dealing with frequency domain is somehow computationally expensive because we need to carry out 2-D Fourier transform and thus it is not easy to apply this technique to large-scale problems. Kumar et al. [101] proposed correlation filters to reduce the computational complexity. They have used Minimum Average Correlation Energy (MACE) Filter [103] to represent one class for each of the $N$ different classes that exist in the face space. Once the filters are designed for each class, they are used to produce the output feature vector for any probe image. One of the benefits of correlation filters is the
graceful degradation they offer because of the integrative nature of the matching operation. If some of the pixels are occluded, they simply do not contribute to the correlation peak. Fig. 2.14 shows an example for a correlation output when the input image is an authentic one (Fig. 2.14-Left) and an impostor (Fig. 2.14-Right).

![Figure 2-14 Correlation outputs for: authentic input (Left) and impostor input (Right).](image)

### 2.4.2 Radial Basis Function (RBF)

Radial functions are a class of functions. A typical radial basis function, centred at $c_j$ and of width or radius $r_j$, is the Gaussian

$$h_j(x) = \exp\left(-\frac{(x-c_j)^2}{r_j^2}\right)$$  \hspace{1cm} (2.6)

Some researchers have applied RBF networks to person recognition [53, 54]. Sato et al. [55] described a face recognition system that uses partial face images (for example, eye, nose, and ear image) for input data.

### 2.4.3 Discrete Wavelet Transform

Shams [68] proposed a method for face recognition based on the calculation of the wavelet coefficients. A discrete wavelet transform is performed on two-dimensional signals $f(x, y)$ using the following equations
\[ A = [h^*\lfloor h^* f \rceil_x \downarrow 2]_y \downarrow 2 \]
\[ H = [g^*\lfloor h^* f \rceil_x \downarrow 2]_y \downarrow 2 \]
\[ V = [h^*[g^* f\xi]_x \downarrow 2]_y \downarrow 2 \]
\[ D = [g^*[g^* f\xi]_x \downarrow 2]_y \downarrow 2 \]  

where * denotes the convolution operator and \( \downarrow 2 \) represents down sampling along \( x \) (or \( y \)) direction. \( g \) and \( h \) are high-pass and low-pass filters, \( A \) is wavelet approximation of the signal and \( H, V, \) and \( D \) are horizontal, vertical, and diagonal wavelet coefficients [69]. For each of \( A, H, V, \) and \( D \) wavelet coefficients, related mean data is subtracted from the coefficients and is projected into the corresponding subspace. The distances between projected input and class prototypes in each of subspaces are computed. The recognition is done based on the results from each of the subspaces separately or using a combination of the results of all subspaces.

### 2.4.4 Facial Asymmetry

Another interesting parameter in face recognition systems is discussed in [109, 110, 111], where facial asymmetry in face recognition is explored and investigated. Mitra et al. [108] proposed facial asymmetry measures in the frequency domain. They have suggested a connection between facial asymmetry and Fourier transform components. Facial asymmetry can be caused either by external factors such as expression changes, viewing orientation and lighting direction, or by internal factors such as growth, injury, and age related changes. The former can be controlled to a large extent, while the latter is more challenging and interesting, and it is known that the recognition rate changes dramatically by such changes.

### 2.4.5 Infra-Red Imagery

Recently, researchers have investigated the use of near infrared (near IR) imagery for face recognition with good results. Near IR imagery like visible imagery is formed from reflected radiation, and the imaging process still requires an external source of illumination [85, 86]. However, the advantage gained from working with this type of sources is that the eye is not sensitive in this range, and illumination can be used in a more flexible and possibly covert way. Some of the research efforts in thermal face recognition use the thermal IR band as a way to see in the dark or reduce
the harmful effect of light variability. Further, attempts have been made to fuse the visible and IR modalities to increase the performance of face recognition [87, 88, 89, 90].

Buddharaju et al. [94] presented an approach to the problem of thermal facial recognition to fully utilize the potential of the thermal IR band. The approach consists of statistical face segmentation and a physiological feature extraction algorithm tailored to thermal phenomenology. The physiological vector is formed from the thermal imprint of the facial vascular network. By this, they attempt to develop a face recognition system using physiological information on the face. The convective heat transfer effect from the flow of "hot" arterial blood in superficial vessels creates characteristics thermal imprints. The pattern of the underlying blood vessels and the corresponding thermal imprints is quite complex.

Figure 2-15 Generic map of superficial blood vessels on the face. (a) Overview of an arterial network. (b) Overview of a venous network. (c) Arteries and veins together underneath the surface of the facial skin [94].

Figure 2-16 Vascular network extraction [94].
As it was argued, this complex pattern is distinctive to each individual and can serve as a useful biometric signature. The thermal imprints are different from one individual to another because of the variable absorption from different fat padding (skinny faces versus puffy faces) and variable heat conductance from different skin complexion (dark skin is less conductive) [91, 92, 93]. Fig. 2.15 shows generic map of superficial blood vessels on the face and Fig. 2.16 shows vascular network extraction result [94].
2.5 Face Recognition for Limited Memory Applications

In a parallel direction of research, limited memory and fast algorithms for face recognition have received some attention from the research community during the last five years. It is noticed that the cell phones have become an integrated part of our lives, where they do more than just enable communication between people. Crucial information is stored in these devices. This information includes contact details, bank account statements, and other private information that need to be secured to guarantee a high level of protection, especially in the cases where the phone is stolen or lost. Biometrics authentication can play a very important role in protecting the information and preventing the access of unauthorised people to such significant information [32]. Face recognition is one of these biometric authentication techniques that is very suitable for such applications. Since most of the mobile phones come with an integrated camera, this makes them very convenient to implement such secure systems.

One of the disadvantages of using biometric recognition is that there is a considerable amount of calculation involved. The challenge is more appreciated when we know that the available processors in the market that are suited for mobile phones contain CPUs that range between 100MHz to 500 MHz. That means the processor speed is very limited. Moreover, the memory of these CPUs is also limited. So we need to consider this important issue in our implementation.

To the best of our knowledge, not much work has been done in this field. Although the face recognition research field has been investigated for more than 30 years, nowadays this area deserves more attention. Note that the mobile technology did not find interesting applications in our lives before a few years ago. This issue is related mainly to the development of the technology of the mobile phone and wireless communication systems implementation. Multimedia applications have only recently found their position in the global market.

Of the published research work on face recognition on mobile phones, Yang et al. [119] proposed a system that is implemented on PDA. This system utilizes the LDA technique. Ng. et al. [8] proposed another system that copes with illumination and pose variations using boosting algorithm to determine which pose variations of the face are challenging and to bootstrap them into the filter synthesis. They have used
Minimum Average Correlation Energy (MACE) filters. Lee et al. [7] presented a face authentication algorithm that uses the features chosen by Genetic Algorithms as an input vector to a support vector machine (SVM). In [120], authors proposed a Digital Signal Processor (DSP)-based platform for face recognition. These processors are supposed to be optimized for processing audio, video, image, and voice signals in power constrained applications. They have used an eigenface algorithm in this platform. Sim et al. [121] introduced a simple memory-based algorithm for face recognition, termed ARENA. The ARENA algorithm employs reduced-resolution images (16×16) which are created by simply averaging over non-overlapping rectangular regions in the image. The distance from the query image to each of the stored images in the database is computed, and the label of the best match is returned. Annesley et al. [122] investigated MPEG-7 colour descriptors in visual surveillance retrieval problems. A set of image sequences of pedestrians entering and leaving a room, viewed by two cameras, are used to create a test set. The problem posed is the correct identification of other sequences showing the same person as contained in an example image. They have used colour descriptors from the MPEG-7 standard, including dominant colour, colour layout, colour structure, and scalable colour. Moreover, Bourlai et al. [160, 161, 162] proposed an implementation of a face verification system on a smart card using the client specific linear discriminant analysis technique (CSLDA). In their work, they have studied the trade-off between the performance and the computational complexity. They have shown that the use of less than 8 bits per pixel gray-scale image resolution does not necessarily result in a degradation of system performance. Further, they have shown that the image resolution can be reduced to a certain level without degrading the system performance. Note that, the client specific linear discriminant analysis technique [163] combines face representation and decision making into a single step, which offers some benefits and advantages over other known methods.
2.6 Conclusion

To conclude this chapter, machine recognition of faces has emerged as an active research area spanning different disciplines such as image processing, pattern recognition, computer vision, computer graphics, and neural networks. The heavy potential use of the face as an identification and verification tool makes it one of the most important and powerful biometrics means. During the last 30 years, significant progress has been achieved on various aspects of face recognition, for example: segmentation and feature extraction. Although many face recognition techniques have been proposed and have shown significant promise, robust face recognition is still a challenge. It might seem that face recognition has reached a mature level and acceptable performance, but actually there are many associated problems that need to be solved. Some of these problems are the ability to recognize faces in harsh environments such as variable illumination and the different poses of the face in the image under test. Moreover, the aging problem is still not fully investigated. In general, it is safe to claim that the appearance-based approaches seem to be dominating currently.

Further, with rapid advancement in the field of digital communication and wireless transmission, there is a great potential for important and vital applications for face recognition systems utilizing portable and small devices. The constraints of the high speed and low memory requirements for such systems should challenge the researchers to come up with new techniques or maybe improve the existing ones, in order to cope with such constraints and difficult conditions. As we have seen in chapter 2, and to the best of our knowledge, there are few published works that propose solution for devices of limited memory constraints. As mentioned earlier, the typical face recognition algorithms that found great applications in different fields (banks and airports) may not be executable in devices that are memory-constrained, with processor speed of no more than 100-500 MHz. This type of systems imposes a great challenge on the memory requirement of the data to be sent. Chapter 4 of this thesis will discuss this issue and try to propose solutions to deal with such situations. We try to reduce the size of the space needed to represent the face features, while keeping the performance rate as high as possible. Further, our proposed system offers low storage facility for the limited memory devices.
Furthermore, results in the literature strongly suggest that there is a benefit in component-based approaches. Component-based approach will help in compensating for pose changes and other disadvantages of holistic methods. As we have seen in this chapter, although component-based approach has attracted researchers’ interest recently, the works presented in this field still suffer from some shortages. To mention some:

- Large processing and computational time requirement
- Requiring many pre-processing steps.
- Large image size used
- Selection of the components in done manually and most of them start with components located around pre-selected points on the face.

Chapter 5 of the thesis will present a novel study of the component-based method that will try to address some of the aforementioned issues.

In brief, it might seem that the face recognition has reached a mature level and acceptable performance, but actually there are many associated problems remain to be solved.
Chapter 3
Pattern Recognition Methods

The purpose of pattern recognition is to determine to which category or class a given sample belongs. Through an observation or measurement process, we obtain a set of numbers which make up the observation vector, which serves as the input to a decision rule by which we assign the sample to one of the given classes. It is important to find and know beforehand which of the features taken will work best, and the implications of high dimensionality. We need to manage the tradeoffs and predict how the system can work when we generalize it to new patterns. The focus is usually on finding the statistical properties of the pattern, where it is expressed in probability densities. The success of the classifier is determined by how well it can quantify near and far apart the query pattern [23, 99].

Since we are interested in the methods and applications of pattern recognition and classification, we will present a brief study of two famous approaches that are extensively used in face recognition field. The two approaches are used as part of the analysis and tools in our work, as will be seen in subsequent chapters. The two famous approaches to find linear transformations are: the Principal Component Analysis (PCA), which seeks best representation of the data, and the Linear Discriminant Analysis (LDA) that best separates the data classes, both in a least squares sense. The PCA is discussed in section 3.1, while the LDA is discussed in section 3.2.

3.1 Principal Component Analysis

3.1.1 Mathematical Background
Given $n$ $d$-dimensional vectors $x_1, x_2, \ldots, x_n$, we can represent any vector $x_k$ from this set of vectors by projecting $x_k$ onto a line running through the mean of the vectors $\mathbf{m}$, by

$$x_k = \mathbf{m} + a_k \mathbf{e} \quad (3.1)$$
where

\[ m = \frac{1}{n} \sum_{k=1}^{n} x_k \]  

(3.2)

and \( e \) is a unit vector in the direction of the line, and \( a_k \) corresponds to the distance of any vector \( x_k \) from the mean \( m \). Now we can find an optimal set of coefficients \( a_k \) by minimizing the squared-error criterion function

\[
J_1(a_1, \ldots, a_n, e) = \sum_{k=1}^{n} \| (m + a_k e) - x_k \|^2 = \sum_{k=1}^{n} \| a_k e - (x_k - m) \|^2
\]

(3.3)

\[
= \sum_{k=1}^{n} a_k^2 \| e \|^2 - 2 \sum_{k=1}^{n} a_k e^T (x_k - m) + \sum_{k=1}^{n} \| x_k - m \|^2
\]

and since \( \| e \| = 1 \) by partially differentiating with respect to \( a_k \), setting the derivative to zero, we obtain

\[ a_k = e^T (x_k - m) \]  

(3.4)

In other words, we obtain least-squares solutions by projecting the vector \( x_k \) onto the line in the direction of \( e \) that passes through the mean of the vectors. In order to find the best direction of \( e \), we need to define the scatter matrix \( S \). The scatter matrix \( S \) is given by

\[ S = \sum_{k=1}^{n} (x_k - m)(x_k - m)^T \]  

(3.5)

Substituting Eq. (3.4) into Eq. (3.3), we obtain
\[ J_1(\mathbf{e}) = \sum_{k=1}^{n} a_k^2 - 2 \sum_{k=1}^{n} a_k^2 + \sum_{k=1}^{n} \|x_k - \mathbf{m}\|^2 \]

\[ = - \sum_{k=1}^{n} [\mathbf{e}^T (x_k - \mathbf{m})]^2 + \sum_{k=1}^{n} \|x_k - \mathbf{m}\|^2 \]

\[ = - \sum_{k=1}^{n} \mathbf{e}^T (x_k - \mathbf{m})(x_k - \mathbf{m})^T \mathbf{e} + \sum_{k=1}^{n} \|x_k - \mathbf{m}\|^2 \]

\[ = -\mathbf{e}^T \mathbf{S} \mathbf{e} + \sum_{k=1}^{n} \|x_k - \mathbf{m}\|^2 \]  \hspace{1cm} (3.6)

It is clear from Eq. (3.6) that the vector \( \mathbf{e} \) that minimizes \( J_1 \) also maximizes \( \mathbf{e}^T \mathbf{S} \mathbf{e} \). If the method of Lagrange multipliers is used to maximize \( \mathbf{e}^T \mathbf{S} \mathbf{e} \) subject to the constraint that \( \|\mathbf{e}\|=1 \), then let \( \lambda \) be the undetermined multiplier to obtain Eq. (3.7)

\[ \mathbf{u} = \mathbf{e}^T \mathbf{S} \mathbf{e} - \lambda \mathbf{e}^T \mathbf{e} \]  \hspace{1cm} (3.7)

Differentiating with respect to \( \mathbf{e} \) yields

\[ \frac{\partial \mathbf{u}}{\partial \mathbf{e}} = 2\mathbf{S} \mathbf{e} - 2\lambda \mathbf{e} \]  \hspace{1cm} (3.8)

Setting this gradient to zero, then \( \mathbf{e} \) must be an eigenvector of the scatter matrix

\[ \mathbf{S} \mathbf{e} = \lambda \mathbf{e} \]  \hspace{1cm} (3.9)

In particular, because \( \mathbf{e}^T \mathbf{S} \mathbf{e} = \lambda \mathbf{e}^T \mathbf{e} = \lambda \), it follows that to maximize \( \mathbf{e}^T \mathbf{S} \mathbf{e} \), we need to select the eigenvector corresponding to the largest eigenvalue of the scatter matrix.

This result can be extended from a one-dimensional projection to a \( d' \)-dimensional projection. Instead of Eq.(3.1), we can write

\[ x_k = \mathbf{m} + \sum_{i=1}^{d'} a_{ki} \mathbf{e}_i \]  \hspace{1cm} (3.10)
where \( d' \leq d \). In this case, the vectors \( \Theta_1, \ldots, \Theta_{d'} \) are the \( d' \) eigenvectors of the scatter matrix having the largest eigenvalues. The coefficients \( a_{ki} \) in Eq. (3.10) are called the \textit{principal components}.

### 3.1.2 PCA for Face Recognition

Images of faces, being similar in overall configuration, will not be randomly distributed in a huge image space and thus can be described by a relatively low dimensional subspace. Principal Component Analysis (PCA) technique, or the "Eigenfaces" method, emphasizes the significant local and global "features" of the face image. Such features may or may not be directly related to our intuitive notion of face features such as the eyes, nose, lips, and hair. The approach was basically suggested by Turk and Pentland [10]. The main idea of the principal component analysis is to find the vectors that best account for the distribution of face images within the entire image space.

It depends on extracting the information contained in an image of a face and to somehow capture the variation in a collection of face images, and use this information to encode and compare individual face images. This is done through finding the principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images. The eigenvectors are ordered, each one accounting for a different amount of the variation among the face images. These eigenvectors can be thought of as a set of features that together characterize the variation between face images. Each image contributes more or less to each eigenvector, so that the eigenvector is displayed as a sort of ghostly face which is called an eigenface.

If the face image \( I(x,y) \) is a two-dimensional \( N \) by \( N \) array of intensity values, then an image may be reconsidered as a vector of dimension \( N^2 \), Fig. 3.1.
If the training set of face images is $\Gamma_1, \Gamma_2, \Gamma_3, \ldots, \Gamma_M$. Then, the average face of the set is defined by

$$\Psi = \frac{1}{M} \sum_{i=1}^{M} \Gamma_i \quad (3.11)$$

Each face differs from the average by the vector $\Phi_i = \Gamma_i - \Psi$. This set of vectors is subject to principal component analysis. These components are the eigenfaces of the covariance matrix

$$C = AA^T \quad (3.12)$$

where the matrix $A = [\Phi_1, \Phi_2, \ldots, \Phi_M]$. The matrix $C$, however, is $N^2$ by $N^2$, and determining the $N^2$ eigenvectors is not reliable.

Turk and Pentland suggested for solving the $N^2$-dimensional eigenvectors by first solving for the eigenvectors of $M$ by $M$ matrix and then taking appropriate transformation or calculation to get the eigenvectors of matrix $C$. If $v_i$ is the eigenvectors of $A^T A$, which is of dimension of $M$, then $u_i$, the eigenvectors of $C$ is given by

$$u_i = \frac{1}{\sqrt{\lambda_i}} A v_i \quad (3.13)$$
where \( \lambda_i \) is the eigenvalue corresponding to the eigenvector \( v_i \). Now, \( u_i \) is a vector of length \( N^2 \) that describes an \( N \) by \( N \) image, and it is a linear combination of the original training face images. In practice, a smaller \( M' \) (with the largest associated eigenvalues) is sufficient for identification. The eigenfaces span an \( M' \)-dimensional subspace of the original image space \( N^2 \).

Now, a face image \( \Gamma \) is transformed into its eigenface components by the following equation,

\[
\omega_k = u_k^T (\Gamma - \Psi)
\]  
(3.14)

for \( k = 1, \ldots, M' \). The weights form a vector \( \Omega^T = [\omega_1, \omega_2, \ldots, \omega_{M'}] \) that describes the contribution of each eigenface in representing the input face image, treating the eigenfaces as a basis set for face images.

Turk and Pentland [10] have conducted several experiments to assess the performance of eigenface technique for face recognition. They have created a database of images digitized under controlled conditions. Sixteen subjects were digitized at combinations of three head orientations, three head sizes or scales, and three lighting conditions. In the classification process, where they have reported different classification rates for the different experiments that have been conducted, they have used seven eigenfaces corresponding to the largest seven eigenvalues.
3.2 Linear Discriminant Analysis

3.2.1 Mathematical Background

Linear Discriminant Analysis (LDA) has been recognized as a powerful method for feature extraction and has been exhaustively used in face recognition due to its optimal performance and time-efficient matching for multi-class face recognition. LDA emphasizes variations among different classes while the variations of the same class are de-emphasized [12, 23, 99, 21, 123].

Linear Discriminant Analysis (LDA), or Fisher Linear Discriminant (FLD) as it is called, is a well known method of finding best components for classification [12, 21]. The number of the components used is usually smaller than the original dimensionality of data vectors, which makes it an adequate method for data reduction. The reduced data space spanned by a certain small number of the components is called a subspace. Although in the early work of face recognition field, people have used the PCA for classification, the LDA has proven its efficiency over the PCA. The principal components are not necessarily useful for data classification. While the PCA seeks the components which are efficient for data representation, the Fisher linear discriminant finds the components which are effective for classification [10, 23, 99, 21, 12, 44, 123].

The efficiency of Fisher's Linear Discriminant method comes from the sense that the method finds a direction on which data are well class-wise clustered. Considering two class separation problem, assume that a set of \( n \) \( N \)-dimensional samples \( \{x_1, \ldots, x_n\} \) is given. The set is partitioned into two subsets of vectors \( X_1 \) and \( X_2 \), each of which corresponds to a class label \( \omega_1 \) or \( \omega_2 \), respectively. Each subset \( X_i \) consists of \( n_i \) sample vectors. A linear combination of the components of \( x \) is defined with \( W \) such that

\[
y = W^T x
\]

where \( y \) is an output vector. The set of \( M \) output vectors \( Y_1, \ldots, Y_M \) can be divided into the two subsets \( Y_1 \) and \( Y_2 \) which correspond to \( X_1 \) and \( X_2 \), respectively. From geometrical point of view, \( y_i \) is the projection of the \( x_i \) onto a line in the direction of \( W \).
To find the optimal direction of $W$ which yields accurate classification of the vectors, we should define separation measures of the transformed (or projected) vectors. A measure of the separation between the projected vectors of the two different classes $\omega_1$ and $\omega_2$, is defined as the difference of the class means. If $m_i$ is given as the $N$-dimensional class mean vector of the data vectors $x$ in $X_i$ by

$$m_i = \frac{1}{n_i} \sum_{x \in X_i} x$$  \hfill (3.16)$$

then the class mean of the projected data vectors is

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{x \in X_i} w^T x = w^T m_i$$  \hfill (3.17)$$

That is, the class mean of the projected data is represented as the simple projection of $m_i$ by $w$.

The distance of the class means of the projected data as a measure of the separation of the different class vectors is defined by

$$|\tilde{m}_1 - \tilde{m}_2| = \left| w^T (m_1 - m_2) \right|$$  \hfill (3.18)$$

This difference must be maximized in order to assume the best classification. The difference of the class means should be relatively maximized to a certain measure of the standard deviations of each class. The scatter of the projected samples within the class $\omega_i$ is defined by

$$\tilde{\sigma}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2 = \sum_{x \in X_i} (w^T x - w^T m_i)^2.$$  \hfill (3.19)$$

Then, the summation $(1/n)(\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2)$ expresses the variance of the mutual data. $\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2$ is called the total within-class scatter of the projected samples [23, 123].

The criterion function of the Fisher linear discriminant with respect to $W$ is given by

$$J(W) = \frac{\left|\tilde{m}_1 - \tilde{m}_2\right|^2}{\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2}$$  \hfill (3.20)$$

The Fisher linear discriminant finds the direction of $W$ which maximizes $J(W)$.

By defining the scatter matrices of within-class and between-class, the Fisher's objective function $J(.)$ can be explicitly represented with $W$. 

41
The scatter matrix $S_i$ is defined as
\[ S_i = \sum_{x \in X_i} (x - m_i)(x - m_i)^T \]  
(3.21)

then the within-class scatter matrix $S_w$ is defined as the sum of the scatter matrices by
\[ S_w = S_1 + S_2 \]  
(3.22)

The scatter matrix of the between-class is similarly defined by
\[ S_B = (m_1 - m_2)(m_1 - m_2)^T \]  
(3.23)

The total within-class scatter of the projected samples in Eq. (3.19) can be written explicitly with $W$ by
\[ S_f + S_l = \sum_{i=1}^{2} \sum_{x \in X_i} w^T (x - m_i)(x - m_i)^T w \]  
(3.24)

Similarly, the measure of the separations of the projected means in Eq. (3.18) is also represented explicitly with $W$ as
\[ (\tilde{m}_1 - \tilde{m}_2)^2 = (w^T m_1 - w^T m_2)^2 \]  
(3.25)

\[ = w^T (m_1 - m_2)(m_1 - m_2)^T w \]

\[ = w^T S_B w \]

The within-class scatter matrix can be singular when the number of sample vectors is smaller than the dimensionality of the vectors, i.e. $n < N$. The Fisher's criterion function $J(.)$ can be re-written with the terms of the scatter matrices $S_B$ and $S_w$ as
\[ J(w) = \frac{w^T S_B w}{w^T S_w w} \]  
(3.26)

This formula is the well-known generalized Rayleigh quotient in mathematical physics [23, 123]. Clearly, the function $W$ which maximises $J(.)$ is the solution of the general eigenvalue problem such that
\[ S_B w = \lambda S_w w \]  
(3.27)
where \( \lambda \) is the constant called eigenvalue. If the within-class scatter matrix \( \mathbf{S}_w \) is non-singular, the generalized eigenvalue problem can be replaced with

\[
\mathbf{S}_w^{-1}\mathbf{S}_b \mathbf{w} = \lambda \mathbf{w}
\]

(3.28)

The solution \( \mathbf{w} \) is obtained by solving a conventional eigenvalue problem.

For the two-class separation case, the solution of the function \( \mathbf{w} \) that optimizes the eigenvalue problem is equivalent to

\[
\mathbf{w} = \mathbf{S}_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)
\]

(3.29)

This solution maximizes the Fisher's criterion by maximizing the ratio of the between-class scatter to the within-class scatter. It is worth mentioning that the Fisher's linear discriminant function is identical to the Bayes' optimal discriminant function when the classes have equal covariance matrices. In other word, the Fisher's linear discriminant function has exactly the same solution as the Bayes’ discriminant function [23, 123].

The analysis developed above can be extended and generalized to the \( c \)-class case. For the \( c \)-class problem, a projection from a \( N \)-dimensional space to a \((c - 1)\)-dimensional space is performed. Thus, the generalization of Fisher's linear discriminant involves \( c-1 \) discriminant functions.

More precisely, assume that a data matrix \( \chi = \{x_1, x_2, \ldots, x_M\} \in \mathbb{R}^{N \times M} \) is given, where \( x_i \in \mathbb{R}^N \) is a \( N \)-dimensional column vector representing the face image. Each vector belongs to one of \( C \) object classes \( \{\chi_1, \chi_2, \ldots, \chi_C\} \). Classical discriminant analysis aims to derive a transformation \( \mathbf{W} \in \mathbb{R}^{N \times n} \) (\( n \leq N \)) which maps a vector \( x \) to \( \mathbf{y} = \mathbf{W}^T x \), \( \mathbf{y} \in \mathbb{R}^n \) such that the transformed data have maximum separation between classes and minimum separation within classes. The between-class and within-class scatter matrices in LDA for the \( c \)-class case are given respectively by

\[
\mathbf{S}_B = \sum_{i=1}^{C} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T
\]

(3.30)

\[
\mathbf{S}_W = \sum_{i=1}^{C} \sum_{x \in \chi_i} (x - \mathbf{m}_i)(x - \mathbf{m}_i)^T
\]

(3.31)
where \( m_i \) denotes the class mean, \( m \) is the global mean of the entire sample set and \( n_i \) denotes the number of samples in class \( c \).

3.2.2 LDA for Face Recognition

Etemad et al. [12] proposed the famous LDA-based feature extraction for face recognition. They have explained that PCA provides the features that capture the main directions along which the face images differ the most, but the within class scatter of the feature points is reduced. However, the LDA uses the information of each class (individual) and finds the representation in a way that emphasizes the variations between the various classes and de-emphasizes the variations between images of the same class. The major problem with the LDA method is the singularity of the within class scatter matrix.

Given the between class scatter matrix \( S_B \) and the within class scatter matrix \( S_W \) for the face space, the attempt is to solve the generalised eigenvalue problem

\[
S_B u_i = \lambda_i S_W u_i \tag{3.32}
\]

From this equation the \( \lambda_i \)'s can be computed as the roots of the characteristic polynomial

\[
|S_B - \lambda_i S_W| = 0 \tag{3.33}
\]

and then the \( u_i \)'s can be obtained by solving

\[
(S_B - \lambda_i S_W) u_i = 0 \tag{3.34}
\]

only for the selected largest eigenvector.

Belhumeur et al. [21] has proposed a solution for the problem of singularity of the within class scatter matrix by first projecting the image into a lower dimensional space using the PCA technique and then applying the LDA. This will ensure that the within class scatter matrix is non-singular.
The optimal projection $\mathbf{w}_{opt}$ that maximises the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples if found by the following

$$\mathbf{W}_{opt}^T = \mathbf{W}_{fld}^T \mathbf{W}_{pca}^T$$  \hspace{1cm} (3.35)

where

$$\mathbf{W}_{pca} = \arg \max_{\mathbf{w}} \left| \mathbf{W}^T \mathbf{S}_T \mathbf{W} \right|$$  \hspace{1cm} (3.36)

$$\mathbf{W}_{fld} = \arg \max_{\mathbf{w}} \left| \frac{\mathbf{W}^T \mathbf{W}_{pca}^T \mathbf{S}_B \mathbf{W}_{pca} \mathbf{W}}{\mathbf{W}^T \mathbf{W}_{pca}^T \mathbf{S}_W \mathbf{W}_{pca} \mathbf{W}} \right|$$  \hspace{1cm} (3.37)

where $\mathbf{S}_T$ is the total scatter matrix. The use of PCA for dimension reduction in the Fisherface method has two purposes. It solves the small sample size problem, that is when within class scatter matrix is singular, the PCA is used to reduce the dimension such that the within class scatter matrix becomes non-singular. The second purpose is that the PCA reduces the computational complexity.

LDA has been extensively used in the last few years, where many forms or ideas that are related to LDA have been introduced, in order to overcome some of the problems associated with it and to ensure better performance [77, 100].
3.3 Conclusion

This chapter presented and discussed two fundamental mathematical tools for statistical decision-making processes in pattern recognition: the PCA and the LDA. In practice, the dimensional space of an image is too large to allow robust and fast recognition. A common way to attempt to solve this problem is to use dimensionality reduction technique such as the PCA and the LDA. The PCA can be used to find a subspace whose basis vectors correspond to the maximum-variance directions in the original space. On the other hand, LDA searches for those vectors in the underlying space that best discriminate among classes. As a result, these two tools find great applications in the field of object recognition in general and in the field of face recognition in particular.
Chapter 4
Limited Memory and High Speed Face Recognition Using Binarized Eigenphases

4.1 Introduction

In the last few years, researchers in the area of face recognition have proposed many numerous techniques that achieve high recognition rate [1, 31, 33]. Despite the significant advances in face recognition technology, it has yet to achieve levels of accuracy required for many commercial and industrial applications. With the increased commercial interest in portable devices and with the advanced real-time face recognition systems, the need for more cost effective and low power implementation of these systems has increased. The typical face recognition algorithms that found great applications in different fields (banks and airports) may not be executable in devices that are memory-constrained, with processor speed of no more than 100-500 MHz. For example, in the surveillance systems where the wireless domain for data transmission is adopted, the system imposes a great challenge on the memory requirement of the data to be sent. Consequently, these devices need special treatment and more investigation.

One of the main challenges of face recognition algorithms is the considerable amount of calculations and computations involved in the process, and eventually this will slow down the system significantly. To the best of our knowledge and as we have shown in chapter 2, there are little published works that propose solution for devices of limited memory constraints [7, 8, 9, 120, 121, 131, 160].

Most of the appearance-based methods for face recognition that deal with the face images as a whole, depend on calculating the eigenvalues and the eigenvectors of a system representing the face space [10-13, 15]. The calculations related to (and consequently the time required for) finding the eigenvectors and eigenvalues of the covariance matrix of the system is relatively huge. For instance, in the case of low memory devices, we can not just apply the principal component analysis (PCA) or any
other known face recognition method such as linear component analysis (LDA) to the face images directly.

In this chapter, we propose a novel technique for face recognition for systems with a limited memory. Thus, we describe a new approach for face recognition system that can be implemented and loaded on small and portable devices. The main aim beyond this process is to dramatically reduce the size of the space needed to represent the face features, while keeping the performance rate as high as possible.

It will be shown that the new technique improves the performance of the recognition rate when compared to the MPEG-7 FFD vector method (explained in section 4.2), the direct eigenphase implementation method to the face space [15], and the PCA method.

The novelty of this technique comes from the following ideas:

a) The face space is not represented by the images of the database in use (as it is done usually); rather it is represented by the Fourier transform of the covariance matrix constructed from the MPEG-7 Fourier Feature Descriptor (FFD) vectors of these images as a first step. By using the MPEG-7 face descriptor, we can end up with a compact vector of small memory size, and consequently, the constraints imposed by the limited memory systems are overcome.

b) As Oppenheim et al. showed [19] that the phase angle of the Fourier transform retains most of the information about the image, we will represent our system with the phases of the Fourier transform of the covariance matrix constructed from the FFD vectors. The phase component acts to position the bright and dark spots in the image in order to form regions that are recognizable by a viewer. Similarly, the phase component retains the most important information about the system of FFD vectors in the covariance matrix.

c) In digital image processing, we can find the important features of an image through enhancing the high frequency components of that image. Since this step is implemented on a set of vectors in the frequency domain, we can look at our system as a matrix that has some elements of high frequency and others of low frequency. The enhancing process of these high frequency components can be done by sharpening the system. This operation can be done by a binarization process, as will be shown in section 4.3.
d) Now, we seek the transformation that best describes the space in a way that helps us in further reducing the dimensionality of this system. In the last step of the proposed system, we apply the PCA on the final resultant matrix.

The details of this method are presented in the subsequent subsections.

4.2 MPEG-7 Fourier Feature Descriptor

The MPEG-7 objective is to describe the content of multimedia data so that it can be efficiently searched, accessed, transformed or adapted for use by any device and to support different applications [16]. MPEG-7 is very flexible where improved algorithms can replace previous ones and therefore it is not frozen in time. In addition, many face descriptors for MPEG-7 have been proposed for face retrieval in video streams. For more details about MPEG-7, the reader may refer to the appendix.

MPEG-7 uses the FFD vector to represent the facial feature of an image. The descriptor represents the projection of a face vector onto a set of basis vectors which span the space of possible face vectors. The Face Recognition feature set is extracted from normalized face images each of size 56×46. The FFD vector represents the facial features of an image using a small single vector. This small vector is derived from two feature vectors of a normalized face image; one is a Fourier Spectrum Vector $x_1^f$, and the other is a Multi-block Fourier Amplitude Vector $x_2^f$. Fig. 4.1 shows the process of extracting the FFD for a single image as explained in [17]. This technique was presented and submitted to MPEG-7 community for possible adoption.
Figure 4-1 The process of Fourier Feature extraction for a single image in the MPEG-7 algorithm; Re [F] and Im [F] denotes the real and imaginary parts of the Fourier Transform of the image, and PCLDA refers to the Linear Discriminant Analysis using Principal Component.
The quantized elements represented in the FFD vector, $W_f$ is of size 63. From Fig. 4.1 we see that the normalized image of size $56\times46$ goes through parallel processing or stages: a) the image is taken as a whole and its Fourier spectrum vector $x_f^f$ is found. This vector describes the image globally; b) the image is divided into different blocks and the Multi-block Fourier Amplitude Vector $x_f^2$ is found. This $x_f^2$ vector describes the image locally; c) the $x_f^1$ and $x_f^2$ vectors are projected using Linear Discriminant Analysis using Principal Component (PCLDA) to find the normalized vectors $y_f^1$ and $y_f^2$; d) the vectors $y_f^1$ and $y_f^2$ are joined to form one vector where this vector is projected using Linear Discriminant Analysis (LDA); e) finally this resultant vector is quantized to produce a small, efficient, and descriptive vector $W_f$ of size 63.

Now, the FFD vectors of all images in the system will span a new space and of course the elements of the vectors are of different weight of importance. It is very beneficial to investigate and find out how these important elements or features are distributed. Accordingly, one can emphasize the important features and neglect the relatively less important ones. This process can be achieved by transforming these vectors into another domain, namely the frequency domain. Constructing the face recognition system from the FFD vectors in frequency domain is explained in the next subsection.

### 4.3 System Description

The calculation of the eigenvalues and the eigenvectors of a system of face images (called face space) is the major process in most of the appearance-based methods for face recognition. This calculation takes the major part of the processing time.

To overcome this problem, the system of Fig. 4.2 is proposed. In this figure, it can be seen that the face space is not represented directly by the images of the database in use (as it is done usually); rather it is represented by the Fourier transform of the covariance matrix constructed from the MPEG-7 Fourier Feature Descriptor (FFD) vectors of these images as a first step.
Note that the MPEG-7 face descriptor produces a compact vector of small memory size. From the MPEG-7 FFD vectors $W_f_1, W_f_2, \ldots, W_f_M$ which are found for $M$ images, where $M$ is the number of images in the database, one can derive the average FFD vector of the $W_f_1, W_f_2, \ldots, W_f_M$ vectors

$$m = \frac{1}{M} \sum_{i=1}^{M} W_f_i$$

(4.1)

Next, the average vector, $m$, is used to construct the covariance matrix $C_s$ as:

$$C_s = AA^T$$

(4.2)

where $A = [D_1 D_2 \ldots \ldots D_M]$, and $D_i = W_f_i - m$. Eq. (4.2) can be rewritten as

$$C_s = \sum_{i=1}^{M} (W_f_i - m)(W_f_i - m)^T$$

(4.3)

Note that Eq. (4.3) represents the covariance matrix in the spatial domain and Fig. 4.3 shows a plot of this resultant matrix.
As Oppenheim et al. showed [19] that the phase angle of the Fourier transform retains most of the information about the image; we will represent our system with the phases of the Fourier transform of the covariance matrix constructed from the FFD vectors. The phase component acts to position the bright and dark spots in the image in order to form regions that are recognizable by a viewer. Thus, the phase component retains the most important information about the system of FFD vectors in the covariance matrix. If we denote $C_f$ as the covariance matrix in the frequency domain, then $C_f$ is given by

$$C_f = F_{DFT} \cdot C_s \cdot F_{DFT}^{-1}$$

(4.4)

where $F_{DFT}$ is the Fourier transform matrix containing the Fourier basis vectors. $C_f$ is a complex-valued matrix whose elements are represented by magnitude and phase. Finding the frequency domain of a system should provide us with the frequency contents of the system under study, where some of these components are more important than others. The magnitude of $C_f$ matrix is shown in Fig. 4.4, whereas the phase part of $C_f$ is shown in Fig. 4.5.
Figure 4-5 The phase of the Fourier transform of the covariance matrix \( C_f \): Left in Radians, Right in Degrees.

It is worth mentioning that dealing with frequency domain is somehow computationally expensive because we need to carry out 2-D Fourier transform and thus it is not easy to apply this technique to large-scale problems. But since we are implementing the Fourier transform on small-scale system (where the size of the matrix is 63\times63), the computations are reduced dramatically and it costs an extra negligible amount of time. This is one of the advantages of our system, where it utilizes the advantages of the FFD vectors in an intelligent way.

Now, as the angles of the elements of \( C_f \) retain most of the information of the images, we will concentrate on the phase information. Thus, the phases \( \Phi(i,j) \) of the elements of \( C_f \) are extracted. It is well known in digital image processing that some important features of an image can be emphasized through enhancing the high frequency components of that image. We can look at our system as a matrix that has some elements of high frequency and others of low frequency. The enhancing process of these high frequency components can be done by sharpening the system. This operation can be done by a binarization process where the phase of each element of \( C_f \) is replaced by a binary value (either 1 or 0) according to

\[
B\Phi(i,j) = \begin{cases} 
1 & \text{for } \Phi(i,j) \geq T_H \\
0 & \text{otherwise}
\end{cases}
\]

(4.5)

Here \( T_H \) is a threshold value that is used for binarization. The threshold value has been set to the median of the phases of the 3\times3 neighbourhood elements surrounding the targeted phase. More specifically, we will find the following value:
\[ T_H = \text{Median} \{ \Phi(i, j), \Phi(i + 1, j), \Phi(i - 1, j), \Phi(i, j + 1), \Phi(i, j - 1), \]
\[ \Phi(i + 1, j + 1), \Phi(i - 1, j - 1), \Phi(i + 1, j - 1), \Phi(i - 1, j + 1) \} \]

then we compare this value with \( \Phi(i, j) \) to see if it is larger or smaller, and consequently we assign a value of 1 or 0 to the corresponding phase element \( B\Phi(i, j) \). The reason behind choosing a 3x3 size is that it is one of the smallest filters that we can apply since we are looking for the minimum number of computations and operations. This binarization step will help in locating the features of interest (the phases that contribute most of the discriminative information about the system). It is similar to providing the silhouette of the object when finding the binary of an image. By applying this step, we gain three important advantages:

- Low storage: no more than 1 bit/pixel, instead of representing the pixel with 8 bits, or 5 bits as the proposed in [17].
- Simple processing: the algorithms are in most cases much simpler than those applied to grey-level images
- Enhancing the performance of the system, as will be seen in the following subsection.

Fig. 4.6 shows the resultant matrix after the binarization step.

![Figure 4-6 Result after the binarization step.](image)

A further significant improvement of the proposed system is obtained by reducing the dimensionality of the system in order to cope with the system constraints and requirement. Since each feature adds to the computational burden in terms of processing and storage, the application of the PCA at the final step further reduces the dimensionality of our system.
A value of $M'$ ($M'$ is much smaller than $M$) eigenvectors associated with the largest eigenvalues is sufficient for the recognition process. It is found from the experiments that $M'$=15 is enough to represent the system in an efficient way. Fig. 4.7 shows the principal components of the binarized eigenphase matrix ordered by the corresponding eigenvalues from largest to smallest.

![Figure 4-7 Principal components ordered by the corresponding eigenvalues from largest to smallest.](image)

It is clear from Fig. 4.7 that the choice of $M'$=15 is convenient since it takes the effects of the variances that contribute up to 90% of the overall distribution.

Next, the FFD vector ($W_f$) is transformed into its eigenvector components by the following weight equation,

$$\omega_k = u_k^T(W_f - m) \quad (4.7)$$

where $u_i$ is the $i^{th}$ eigenvector and $k = 1, ..., M'$. The weights form a vector $\Omega^T = [\omega_1, \omega_2, ..., \omega_M]$ that describes the contribution of each eigenvector in representing the input *binarized phase* FFD vector, and consequently the $\Omega^T$ represents the corresponding (original) face image.
We have used the Euclidian distance to determine which face class provides the best description of an input face image by finding the face class $k$ that minimizes the Euclidian distance

$$
\varepsilon_k = \| (\Omega - \Omega_k) \|^2
$$

where $\Omega_k$ is a vector describing the $k^{th}$ binarized phase FFD of a face class and it is calculated by averaging the results of the eigenvector representation over a small number of binarized phase FFD vectors of each individual. A face is classified as belonging to class $k$ if the corresponding $\varepsilon_k$ is the minimum among all other $\varepsilon_k$'s.

### 4.4 Experimental Results

Many different experiments were performed to evaluate the performance of the proposed system. We carried out experiments on two independent and different databases, one is the ORL and the other is the xm2vts. Both sets include a number of images for each person, with variations in pose, expression and lighting. The ORL set includes 400 images of 40 different individuals where each individual is represented by 10 images. The system was trained using 5 images for each person from this set. For the xm2vts set, we have used 2360 images for 295 different individuals with each individual represented by 8 different images. These images have been taken at four different sessions, with two shots at each session.

The xm2vts uses a standard protocol, referred to as the Lausanne protocol. This protocol was defined for the task of verification. The features of the observed person are compared with stored features corresponding to claimed identity, and the system decides whether the identity claim is true or false on the basis of a similarity score. The subjects whose features are stored in the system database are called clients, whereas the person who is claiming a false identity is called an imposter. According to the Lausanne protocol, the database is split into three groups: the training group, evaluation group, and the testing group. We have trained our system with the images from the first two sessions (4 images), and used the images from the third session for evaluation (2 images), and finally we have used the images from the fourth session for
testing (2 images). The evaluation set is used to find the *threshold* that determines if a person is accepted or rejected.

The xm2vts database images are taken at different sessions (different days). The experiments on this database test the robustness of the proposed system under the variation in time conditions of the images. Different timing means different hair style, different clothes and different "moods". Fig. 4.8 shows examples of the xm2vts database used for this experiment. The following steps were carried out on both databases:

a) The images were normalized to a size of 56×46.

b) The MPEG-7 algorithm was applied to all images, and the FFD for each one of the images was calculated.

c) The eigenvectors and eigenvalues of the (63×63) covariance matrix were calculated, the $M'$ eigenvectors corresponding to the highest associated eigenvalues is chosen. $M' = 15$ was selected.

d) For each known individual, the class vector $\Omega_k$ was calculated by averaging the pattern (weight) vectors $\Omega$ for the learning images calculated from the original FFD vector of each individual.

e) For each new face image to be identified, its pattern vector $\Omega$ was found and the distances $e_k$ to each known class was calculated. The class vector $\Omega_k$ that has the minimum distance $e_k$ will represent this input face.

The new system achieved 93% recognition rate when applied to xm2vts database, while under the same conditions the MPEG-7 face recognition method achieved 89%.
Figure 4-8 Examples from the xm2vts database.
The other experiment was to test the proposed technique under other different circumstances. The ORL face database is used in this experiment. This database include images with different poses, different illuminations, different expressions (open or closed eyes, smiling or non-smiling), different facial details (glasses or no glasses), and some of them were taken at different times. Examples of the images in the ORL database used are shown in Fig. 4.9. The proposed technique achieved 95.5% correct classification, while under the same conditions the MPEG-7 face recognition method achieved 91.5% correct classification.

![Figure 4-9 Examples of the images from the ORL database.](image)

A summary of the results of the recognition test on both databases is given in Table 4.1. Note that in this table, the results of applying the PCA method and the direct eigenphase implementation method to the face images [15] are also shown. As seen in Table 4.1, the system is giving better performance when applied to ORL. This is due to the conditions of the images of the two databases. The xm2vts database is more challenging than the ORL one, considering the number of classes, the volume of the database, and the period of time the images were taken. All of these conditions affect the overall performance of the system.
Table 4-1 The recognition rates of four different methods.

<table>
<thead>
<tr>
<th></th>
<th>PCA</th>
<th>Eigenphase</th>
<th>MPEG-7</th>
<th>New method</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORL</td>
<td>78%</td>
<td>81%</td>
<td>91.5%</td>
<td>95.5%</td>
</tr>
<tr>
<td>xm2vts</td>
<td>74%</td>
<td>77%</td>
<td>89%</td>
<td>93%</td>
</tr>
</tbody>
</table>

It is known that any recognition system has a sensitivity threshold that must be set appropriately. Setting the threshold too low will result in too many false positives—an innocent person is subjected to further amount of inspection based on resemblance to a person on a watch list. Setting the threshold too high will result in too many false negatives—a terrorist is not recognized due to differences in appearance between the gallery and probe images of the person under suspect.

To achieve both fewer false positives and fewer false negatives is an important target of any new system, and every researcher try to reach this goal. A method to evaluate the potential performance of the recognition system is the Cumulative Match Score curve (CMS). The CMS illustrates the trade-off of true positive versus false positive results. The curve gives the fraction of the time that when someone on the watch list appears in the surveillance image, the system signals an alarm to the operator. The Y-axis is the true positive rate, and the X-axis is the cumulative rank. We can think of the X-axis as the maximum number of images that the system is allowed to report when giving an alarm for a given probe. If the system is allowed to report a larger number of possible matches, the true positive rate generally increases.

The results of the CMS evaluation for the ORL and xm2vts databases are shown in Fig. 4.10 and Fig 4.11, respectively, where the performance statistics are reported as cumulative match scores. Identification is regarded as correct if the true object is in the top Rank \( n \) matches.
Figure 4-10 The cumulative match scores of the four methods for the ORL database.
Figure 4-11 The cumulative match scores of the four methods for the xm2vts database.
We have also assessed the performance of our new technique by comparing between the methods using the receiver operating characteristic (ROC) curve. The ROC curve is a plot of false acceptance rate against false rejection rate for all values of threshold. The value of this threshold determines the number of false acceptances and false rejections. If the threshold is increased, false acceptance decreases and false rejection increases. Similarly, the opposite is true.

The ROC curves for the ORL and xm2vts databases are shown in Fig. 4.12 and Fig. 4.13, respectively. It is clear from these figures that our proposed method outperforms the other methods (MPEG-7, Eigenphase, and PCA).

![Figure 4-12 The ROC curves of the four methods for the ORL database.](image-url)
Figure 4-13 The ROC curves of the four methods for the xm2vts database.
We have also compared between the computation time of our system and the other systems. Table 4.2 demonstrates the amount of time consumed by each method to perform the training process, when the ORL database is used. As the training is a batch process, the time shown represents the time needed for training all the images.

Table 4-2 Amount of time consumed by each method to perform the training process.

<table>
<thead>
<tr>
<th>Method</th>
<th>Time* (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>3.84</td>
</tr>
<tr>
<td>Eigenphase</td>
<td>5.2</td>
</tr>
<tr>
<td>MPEG-7</td>
<td>11.8</td>
</tr>
<tr>
<td>New method</td>
<td>12.1</td>
</tr>
</tbody>
</table>

*Using Pentium 4 – 2.8 GHz – 512M RAM.

Table 4.2 illustrates that the consume time of the new method is comparable to the time consumed by the MPEG-7 method. However, the performance of the new method is better than the corresponding MPEG-7 one, as it has been shown in Table 4.1. Further, the training time consumed by the new method is approximately twice the time consumed by the eigenphase method and approximately three times the corresponding one of the PCA method. But, as it has been shown earlier, the performance of the new method is the best among all. Also, it is important to note that the first block of the system, i.e. the calculation of the FFD vectors takes the most of the time (11.8 minutes). Whereas it takes only 20 seconds to accomplish the other three stages: the Fourier transform of the covariance matrix takes 8 seconds, the binarization step takes 6 seconds, and finally the PCA takes 6 seconds.

Table 4.3 shows the amount of time consumed by each method to perform the verification process against a claimed identity, which is the time for testing a single image. In this table, a significant reduction in testing time is achieved for our technique when compared to other methods. The testing time for the new method is faster 23 times than the eigenphase one, and 16 times than the PCA one. Note that although the MPEG-7 method provides an attractive testing time, the recognition performance is not as good when compared to our method (see Table 4.1).
Table 4-3 Amount of time consumed by each method to perform the verification of a single image.

<table>
<thead>
<tr>
<th>Method</th>
<th>Time* (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>1.6 sec.</td>
</tr>
<tr>
<td>Eigenphase</td>
<td>2.3 sec.</td>
</tr>
<tr>
<td>MPEG-7</td>
<td>0.06 sec.</td>
</tr>
<tr>
<td>New method</td>
<td>0.1 sec.</td>
</tr>
</tbody>
</table>

*Using Pentium 4 – 2.8 GHz – 512M RAM.

In brief, we can conclude that during the training phase of the recognition system the new method is consuming more time when compared to some other methods, however, during the testing phase it provides very fast as well as the best performance.

4.5 Conclusions

Face recognition techniques that utilize small memory size devices and at the same time show robustness in performance are worth more investigation. In this chapter, a novel method is proposed for efficient face recognition that can be implemented in systems that have limited memory capacities and have low speed processors. The new technique exploits the characteristics and advantages of the MPEG-7 FFD vectors, frequency domain, binarization process, and the PCA. The MPEG-7 FFD provides compact and small memory size vectors. The frequency domain provides the frequency contents of the system, where the important contents are emphasized using a binarization process. The binarization process emphasizes the important features of the covariance system through enhancing the high frequency components. Finally, the PCA is applied to the system for further dimensionality reduction. The new system enhances the performance of the system and provides much better recognition rate compared to other known techniques when applied to two independent face images databases. In addition, the performance of the system was assessed using the CMS and the ROC curves. Both curves have shown that our method gives promising results. Finally, our proposed system offers another very important advantage: it provides low storage. No more than 1 bit/pixel is needed, unlike most other systems that deal with grey-level intensity requiring 8 bits for representing one pixel, or 5 bits per pixel as required by the MPEG-7 original system.
Chapter 5
Component-Based Linear Discriminant Analysis

5.1 Introduction

One of the most important steps in face recognition system is the feature extraction of
the face image. As we have seen in chapters 2 and 3, Linear Discriminant Analysis
(LDA) has been recognized as a powerful method for feature extraction and it has
been exhaustively used in face recognition due to its optimal performance and time-
efficient matching for multi-class face recognition. Thus, LDA emphasizes variations
among different classes while the variations of the same class are de-emphasized [12,
21, 23].

In chapter 4, we presented a solution for face recognition systems which is
constrained by computational time and limited amount of memory that is required for
it to represent the face space. As a result, the performance level or the recognition rate
reached a maximum of ~93%. This recognition rate is very acceptable for such
constrained systems since it is a compromise between the performance for the virtue
of faster and more memory efficient systems.

In addition, part of the literature review presented in chapter 2 covered the
component-based approach in face recognition where the researchers treated the face
image at the level of component images [22, 137, 138, 139, 140, 141, 142, 146, 147,
148, 149, 150, 151, 153, 155, 157, 158, 159]. Although there are good amount of
works done in this field, to the best of our knowledge there is no detailed study
relating the number of components required to represent the face image, the size of
these components, and their locations. Also, the works presented in this area have
some problems, as we have seen in there (please refer to sections 1.1 and 2.6).

In this chapter, we will try to tackle the problem of component-based methods in a
different and a new way. We will try to present a study that relates the three factors
(number of components, their sizes, and their locations) in one side, with the
performance of the system in the other side. Further, we will present the results that relate the performance of the system and the corresponding consumed time.

Being more specific, we present component-based LDA scheme where the face image is partitioned into components and each component undergoes LDA transformation. The component-based LDA has proved to be a very efficient method for face representation [22, 155]. Encoding the facial components separately helps in limiting the image variations to each component region, and hence it facilitates an efficient and robust recognition. Another advantage of the LDA encoding at the component level is that it simplifies the related calculations because we deal with parts of the image and not the whole image. In addition, face occlusion can be more easily identified and dealt with for person identification using the component-based LDA. Moreover, the component LDA effectively solves the problems of face retrieval and person identification, since it proposes a vector of small size and of low dimension. The closest work to what we are proposing is presented in [22]. Kim et al. [22] apply component-based LDA onto facial components and offer a good representation for MPEG-7 face description. However, in their study they do not explain how the facial components are chosen. Further, no criteria are mentioned on how to decide the size of component that should be used, or how many components are needed, or the proper location of these components. Moreover, the time consumed by their method is very large as we will demonstrate in our experiments. This problem was efficiently overcome in our new proposed method.

Therefore, finding the optimal component-based LDA face representation is the objective of this chapter. We propose what can be called a “customer need” application. As will be discussed and seen later, the customer will have a range of alternatives or a range of solutions from which he/she can choose what is suitable for him/her and consequently he/she picks the desirable performance versus the corresponding consumed time. We will try to find the best size, number, and the location of partitions (components) necessary to describe the face image efficiently.

We present extensive experimental results of testing and implementing the proposed approach on two independent databases; the Olivetti Research Ltd. (ORL) database and the xm2vts, CVSSP – University of Surrey database, which are the same databases that have been used in the previous chapter.
5.2 Linear Discriminant Analysis

The LDA has been introduced in some details in chapter 3. We will briefly review some concepts that are necessary in this section. The reduced data space spanned by a certain small number of the directions is called a subspace. For the C-class problem, the linear discriminant analysis involves C-1 discriminant functions. Thus, the projection from a N-dimensional space to a (C-1)-dimensional space is performed.

More precisely, assume that a data matrix $X = \{x_1, x_2, ..., x_M\} \in \mathbb{R}^{N \times M}$ is given, where $x_i \in \mathbb{R}^N$ is a N-dimensional column vector representing the face image. Each vector belongs to one of C object classes $\{X_c, X_i\}$. Classical discriminant analysis aims to a transformation $W \in \mathbb{R}^{N \times n} (n \leq N)$ that maps a vector $x$ to $y = W^T x$, $y \in \mathbb{R}^n$ such that the transformed data have maximum separation between classes and minimum separation within classes. The between class and within class scatter matrices in LDA for the C-class case are given respectively by

$$S_B = \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^T$$  \hspace{1cm} (5.1)

$$S_W = \sum_{i=1}^{C} \sum_{x \in X_c} (x - m_i)(x - m_i)^T$$  \hspace{1cm} (5.2)

where $m_i$ denotes the class mean, $m$ is the global mean of the entire sample set and $n_i$ denotes the number of samples in a class $c$. The LDA finds the direction of $W$ that maximizes $J(W)$. The criterion function $J(.)$ can be written with the terms of the scatter matrices $S_B$ and $S_W$ as

$$J(W) = \frac{W^T S_B W}{W^T S_W W}$$  \hspace{1cm} (5.3)

The within-class scatter matrix can be singular when the number of sample vectors is smaller than the dimensionality of the vectors, i.e. $M < N$. The function $W$ that maximizes $J(.)$ is the solution of the generalized eigenvalue problem

$$S_B W = \lambda S_W W$$  \hspace{1cm} (5.4)
where $\lambda$ is the eigenvalue. If the within-class scatter matrix $W$ is non-singular, the generalized eigenvalue problem can be replaced with

$$S_W^{-1}S_BW = \lambda W$$  \hspace{1cm} (5.5)

The solution $W$ is obtained by solving a conventional eigenvalue problem.

### 5.3 System Description

By considering local pieces of the face images (instead of the entire face image) for recognition, we may achieve higher correct recognition rates. The reason behind this relies on the fact that the image variations on the component level is limited when compared to the variations within the entire “or the global” image.

The component LDA offers a number of advantages and can find many applications:

- **Occlusion problem:** when the image is occluded, we can divide the image into a number of blocks in order to exclude the occluded part of that image and consequently the recognition will be based on the other components, as will be shown shortly.

- **MPEG-7 feature descriptor:** the component LDA technique can be a good candidate for the MPEG-7 face feature descriptor, where we can represent the image with a small vector that is efficient and sufficient for face description.

- **Time efficient and memory efficiency applications:** for systems that are constrained by speed and low memory capacity, the component LDA method gives promising results. More precisely, our experiments have shown that the time required to calculate the different scatter matrices for a $56 \times 46$ image size (the original size of the image that we used in our experiments, as will be described in subsequent section) is quite large. Whereas the total time required to find the corresponding matrices of the different components for the component-based LDA method is much less, and consequently it is much faster to deal with a component face descriptor than a holistic face descriptor. Further, since we are dealing with components of smaller sizes than the original image size, the calculation of the different scatter matrices does not require large amounts of memory and the system will not allocate large memory for such calculations. Hence we obtain memory efficient system.
Different sizes of the patches can be chosen for the components. Selecting the right size is a difficult task and could depend on many factors: the number of images in the database, the size of the images, the variations in the input images, and the required identification speed. These factors influence the decision of choosing the optimal component size, number, and location, which are not independent. Fig. 5.1 shows the proposed system for the optimal component-based LDA method. A face image in the database is partitioned into a number \( (L) \) of components with a predefined size. The size and the necessary number of these components are not randomly selected, rather it is chosen based on required processing time, as will be shown later in section 5.5 when we discuss experimental results. The components comprise what are generally considered to be important features (eyes, nose, and mouth). However, the locations of these windows are automatically found through a search of the best number of subimages (components) that can describe and eventually recognise the face using the Sequential Floating Forward Search (SFFS) algorithm [98].

Figure 5-1 The proposed system for the optimal size, number, and locations of component LDA.

Then for each component we find the LDA transformation of it, where we end up with \( L \) transformations. The test (query) image is divided in the same way as we divide the training images. Now, for every component of the test image we find the
distance between its LDA transformation and the corresponding training components.
LDA transformation. At this stage, we have a candidate for each component representing the identity of the test image. Finally, we find the maximum number of the components giving the same candidate and recognise the test image as that person. This is basically a fusion process where we fuse the information coming from each component to produce a single score that represents the combined decisions of the components result. The fusion method that is used is the Voting method. This method outputs a score equal to the number of component classifiers that output scores above their respective thresholds. For example, if we have five classifiers and three of these classifiers give a score belonging to identity A, while the fourth classifier suggests identity B and the fifth classifier give a result in favour of identity C, then based on the voting method the system will decide that the image belongs to identity A.

On the other hand, apparently it is not feasible to try all possible locations of the components since this will result in a combinatorial explosion. One practical approach to address this issue is to use the “floating” search methods [24, 98] to find the best subset from a set of features. Note that the SFFS was found to be one of the best feature selection methods [98, 99]. The proposed system shown in Fig. 5.1 starts with a specific number of components with a pre-defined size and then it uses the SFFS algorithm to find the best location of these components that can give the highest possible recognition rate. If a pre-defined recognition rate is not achieved, the SFFS algorithm increases the number of components and the process is repeated again. The SFFS algorithm includes new best features (here the feature is the component) that when added to the current feature set, the error rate is minimized. Further, the SFFS algorithm excludes the worst feature during the selection process in order to further improve the selection of the feature set.

In mathematical words, given a set of \( M \) training images \( \{ x_1, x_2, \ldots, x_M \} \), then a set of LDA transformation matrices is extracted. First, all the images are partitioned into \( L \) facial components. The image patches of each component are represented in a vector form with the \( k^{th} \) component being denoted as \( \{ c_{f}^{k}, \ldots, c_{M}^{k} \} \). Then, for the \( k^{th} \) facial component, the corresponding LDA transformation matrix \( W^k \) is computed. During testing, the \( L \) vectors \( \{ c^1, \ldots, c^L \} \) that correspond to the facial component
patches are extracted from a face image $x_i$ of the test data set. Next, a set of LDA feature vectors $y = \{y^1, \ldots, y^L\}$ is extracted from the test image $x_i$ using the corresponding LDA transformation matrices as

$$y^k = (W^k)^T c^k, \quad k = 1, \ldots, L$$

Thus as shown in Fig. 5.1, for the component-based LDA, a face image $x_i$ is represented by a set of LDA feature vectors $\{y^1, \ldots, y^L\}$. In addition, we have used the Euclidian distance in finding the minimum distance. The following section elaborates more on the problem of feature selection and presents the SFFS algorithm steps.
5.4 Feature Selection

The feature selection problem has received considerable attention where numerous algorithms have been proposed [95, 96, 97]. The literature covering feature selection is extensive and spread across many fields. For example, the ability to select features from a huge feature set is critical for computer vision. The main goal of feature selection is to select a subset of $d$ features from the given set of $D$ measurements ($d < D$), without significantly degrading the performance of the recognition system. To guarantee optimality of the selected feature set, one needs to account all the possible subsets of $d$ out of $D$. The number of these sets is given by the following formula

$$\binom{D}{d} = \frac{D!}{d!(D-d)!}$$

(5.7)

For instance, to select five features out of 30 available measurements one would require 142506 evaluations for the feature sets.

In general, any data may contain irrelevant or redundant features that affect the data analysis negatively. Domain information analysis is helpful in pruning the data and in identifying the candidate variables. However, in many cases the size and dimensionality of data make it difficult to use available domain information to identify features that discriminate between classes of interest. Assuming that a suitable criterion function (a selected measure) has been chosen to evaluate the effectiveness of feature subsets, feature selection is reduced to a search problem that detects an optimal feature subset based on the selected measure. In short, the problem of feature selection involves the following issues: large number of features, many irrelevant features, many redundant features, and noisy data. Next, we describe in brief two algorithms which are commonly used in feature selection.

5.4.1 Filters

The filter methods are the simplest methods of selecting a subset of features by making a single measurement of the usefulness of each and thresholding to remove the undesirable features. There are different types of filters that can be used for feature selection. The Kullback-Leibler filter [95] estimates how well a feature
separates the data into different classes using Kullback-Leibler distance between histograms of feature values.

Another filter that finds applications in feature selection ranks the features by sorting them in descending order of Chi-square statistics computed from their contingency tables [95]. The contingency tables have one row for every class and the columns correspond to possible values of the feature. Numeric features are represented by histograms, so the columns of the contingency table are the histogram bins. The Chi-square statistic for feature \( j \) is

\[
\chi^2_j = \sum_i \frac{(o_i - e_i)^2}{e_i} \tag{5.8}
\]

where the sum is over all the cells in the contingency table, \( o_i \) stands for the observed value, and \( e_i \) is the expected frequency of items.

### 5.4.2 Wrappers

Wrapper methods are a more complicated approach to select the best subset of features. The user chooses some method of assessing a subset of features, and this method is \textit{wrapped} in an algorithm that searches, by means of including and excluding features, until the user defined measure is maximized. Sequential forward selection (SFS) and sequential backward selection (SBS) are two classic greedy wrappers [99]. SFS is a “bottom up” search that starts with an empty set of features. In the first iteration, the algorithm considers all feature subsets with only one feature. The feature subset with the highest accuracy is used as the basis for the next iteration. In each iteration, the algorithm adds to the basis each feature that was not added before and retains the feature subset that results in the highest estimated performance. The search terminates after a certain required performance is attained or after the accuracy of the current subset cannot be improved anymore by adding any extra feature. SBS algorithm is a “top down” procedure. SBS works in an analogous way to SFS except that it starts from the full set of features and tentatively deleting some features.

However, both the SFS and SBS methods suffer from the “nesting effect” problem. In the SFS method, the features once selected cannot be later discarded,
while in the case of the SBS method the discarded features cannot be re-selected. An attempt to prevent the nesting of feature subsets, another method was developed: the Plus-/-Minus-r (l-r) search method. The main drawback of this method is that there is no theoretical way of predicting the values of l and r to achieve the best performance set. The floating search methods are generalized (l, r) algorithms which involve successive augmentation and depletion processes. However, instead of fixing the values of l and r, we let these values “float”, i.e., to keep them flexibly changing so as to approximate the optimal solution as much as possible. Because of this “floating” characteristic, these methods have been denoted floating search methods.

Consequently, we have Sequential Forward Floating Search (SFFS) method, which is basically a bottom up search procedure that includes new features by means of applying the basic SFS procedure starting from the current feature set, followed by a series of successive conditional exclusion of the worst feature in the newly updated set provided a further improvement can be made to the previous sets. Similarly, we have Sequential Backward Floating Search (SBFS) method, which is a top down search procedure which excludes features by means of applying the basic SBS procedure starting from the current feature set and followed by a series of successive conditional inclusions of the most significant feature from the available features if an improvement can be made to the previous sets.

Floating search methods provide a performance that has been found to be very good compared with other suboptimal search methods and they are computationally much more efficient than the optimal methods [98]. Moreover, it has been found that SFFS method is more efficient [98] than the SBFS method, therefore we have chosen the SFFS as the feature selection algorithm in our proposed technique for face recognition. Briefly, the SFFS procedure can be explained as follows.

Suppose k features have already been selected from the complete set of measurements \( Y = \{y_j | j = 1,2, ..., D\} \) of D available features to form the set \( X_k = \{x_i : 1 \leq i \leq k, x_i \in Y\} \) with the corresponding criterion function \( J(X_k) \). In addition, the values of \( J(X_i) \) for all preceding subsets of size \( i = 1,2, ..., k-1 \), are known and stored.

77
Step 1. Select feature $x_{k+1}$ from the set of available measurements, $Y - X_k$ to form feature set $X_{k+1}$ such that $J(X_k + x_{k+1})$ is minimised with respect to $x_{k+1}$.

Step 2. Find a feature, $\hat{x}_{k+1}$, in $X_{k+1}$ the removal of which will minimise $J(X_{k+1} - \hat{x}_{k+1})$ with respect to $\hat{x}_{k+1}$.

Step 3. If $x_{k+1} = \hat{x}_{k+1}$, then $k = k + 1$ and return to Step 1.

Step 4. If $x_{k+1} \neq \hat{x}_{k+1}$, then $X_k = X_{k+1} - \hat{x}_{k+1}$.

Step 5. If $k = 2$, then $X_k = \hat{x}_k$, and return to Step 1.

Step 6. Find a feature, $\hat{x}_r$, in $\hat{x}_k$, the removal of which will minimise $J(\hat{x}_k - \hat{x}_r)$.

Step 7. If $J(\hat{x}_k - \hat{x}_r) > J(X_{k-1})$, then $X_k = \hat{x}_k$, and return to Step 1.

Step 8. If $J(\hat{x}_k - \hat{x}_r) \leq J(X_{k-1})$, then $X_{k-1} = \hat{x}_k - \hat{x}_r$, $k = k - 1$.

Step 9. If $k = 2$ return to Step 1, otherwise return to Step 6.
5.5 Experimental Results and Analysis

As we have done in chapter 4, we carried out experiments on the two independent and different databases; the ORL and the xm2vts. Recall that both sets include a number of images for each person, with variations in pose, expression and lighting. The ORL set includes 400 images of 40 different individuals where each individual is represented by 10 images. The system was trained using 5 images for each person from this set. For the xm2vts set, we have used 2360 images for the 295 different individuals with each individual represented by 8 different images. These images have been taken at four different sessions, with two shots at each session [127]. As described in chapter 4, the database is split into three groups: the training group, the evaluation group, and the testing group.

To test the robustness of our technique, we have tested our method on the xm2vts database at three different stages. At the first stage we have used 1180 images for the 295 different individuals taking into account only the images from the first shot of the four different sessions. As such, each individual will be represented by 4 different images. In this case, 2 images of each individual were used for training, one for evaluation, and one for testing.

At the second stage, we have tested our technique using the images from the second shot of the four different sessions. Here again two images were taken for training, one for evaluation, and one for testing.

At the final stage, we have used the whole 2360 images for the 295 individuals where we have trained our system with 4 images from the first two sessions, and used the 2 images from the third session for evaluation, and finally we have used the 2 images from the fourth session for testing. Note that the evaluation set is used to find the threshold that determines if a person is accepted or rejected.

The results obtained from the experiments of the three stages were almost identical, as will be seen later from the CMS and ROC curves. The only difference was in the amount of the time consumed for the three stages. However, it is important to notice that we had to increase the number of eigenvectors that are needed to represent the face space when the database is enlarged (doubled in our case).
Figures 5.2 and 5.3 show examples of the images from the xm2vts and the ORL databases used in the evaluation of our system, respectively.

Example of an image and some of its corresponding components taken from different part of the image are shown in Fig. 5.4. These components are candidates for the system upon which the system chooses the best ones that can give the best performance among the others, as it was explained in sections 5.3 and 5.4.

**Figure 5-2** Different images from 4 different sessions (each session with 2 shots) for a certain individual taken from the xm2vts database.

**Figure 5-3** Examples of the images from the ORL databases.
Figure 5-4 A demonstrative example of different components that correspond to a face image.
5.5.1 Occlusion Problem

As a first experiment to the approach that was presented in Section 5.3, 4 components ($L = 4$) each of size 30×30 for the xm2vts database are used as shown in Fig. 5.5. The component-based LDA finds the individual recognition rate of each component, and then the component-based system fuses the information of each component to produce a single score rate that represents the combined decisions based on a voting method. The final combined rate is shown in Table 5.1. It is worth mentioning that although an excellent successful recognition rate is achieved, the processing time is very large (48.7 minutes), when we have used the whole database (2360 images). As such, we reduced the size of the components by 5 pixels, i.e. the new size will be 25×25. As shown in Table 5.1, the processing time is still high (36 minutes) which is still unsatisfactory for many practical implementations. Therefore, we will keep reducing the size of the components and the same time try to keep the recognition rate as high as possible. This will be our target in the next sub-sections.

![Figure 5-5: Four components face image.](image)

<table>
<thead>
<tr>
<th>Comp. size</th>
<th>Component-based LDA success Rate</th>
<th>Time consumed (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30×30</td>
<td>99.5%</td>
<td>48.7</td>
</tr>
<tr>
<td>25×25</td>
<td>99.2%</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 5-1: The results of the experiments related to Fig. 5.5 (4 components).
As a demonstration of the strength of our proposed technique, we have applied it to the occlusion problem in face recognition. If the given face image was occluded, as shown in Fig. 5.6, our approach can be very efficient in recognizing this image. For the quarter-part occluded images applied to the xm2vts database, we have reached an \( \sim 93\% \) successful recognition rate. When we applied our method to the half-part occluded images, we have achieved an \( \sim 80\% \) successful recognition rate.

![Figure 5-6 Original, one-quarter, and one-half occluded image.](image)

### 5.5.2 Optimality Results for Different Number of Components of Size $20 \times 20$

A main objective of our method is to reduce the size of the components as much as possible in order to decrease the amount of calculation needed. As a result, the issue of the component location arises, since for smaller component sizes, empty uncovered regions of the images will be present. Further, these components should cover the most important regions of the face image. We have applied our technique on images using 4 components each of size $20 \times 20$. We have chosen this size because we have obtained a very attractive processing time. Some initial selection of the 4 components that are proposed by the selected algorithm are shown in Fig. 5.7. Table 5.2 (the results correspond to xm2vts database), shows some examples of the results. The recognition rate for an individual component is found to be above 50\%. However, the overall success recognition rate for the whole component analysis approach is above 73\%. Note that the component number shown in Table 5.2 represents a specific case of combined components layout that are generated by the SFFS algorithm.
A closer look at Fig. 5.7 and Table 5.2 indicates that some parts of the face are more discriminating than other parts. This is an expected finding. For instance, this is demonstrated in component “36” where the 4 components cover the central part of the face which describes parts of the eyes, nose and mouth. Component “36” layout
achieves the highest success rate of 71%. Thus, the central part of a face is very descriptive and discriminative and needs to be considered when generating face components.

Table 5-2 Sample results (from the xm2vts database) of the experiments (components) for images in Fig. 5.7.

<table>
<thead>
<tr>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>54.6</td>
<td>73.9</td>
<td>34</td>
<td>52.4</td>
<td>81.4</td>
<td>38</td>
<td>51.2</td>
<td>76.6</td>
</tr>
<tr>
<td>31</td>
<td>54.7</td>
<td></td>
<td>35</td>
<td>55.4</td>
<td></td>
<td>39</td>
<td>56.6</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>63.6</td>
<td></td>
<td>36</td>
<td>71.0</td>
<td></td>
<td>40</td>
<td>58.5</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>55.9</td>
<td></td>
<td>37</td>
<td>63.4</td>
<td></td>
<td>41</td>
<td>63.4</td>
<td></td>
</tr>
</tbody>
</table>

The results of the 4 components experiments indicate that the number of components should be increased in order to have a reliable high recognition rate. Thus, the next experiment increases the number of components to five. Fig. 5.8 shows some initial selection sets of 5 components that were used in our proposed technique. By increasing the number of the components, a higher overall component LDA success rate ~87.8% is achieved, as can be seen from Table 5.3. Note that the final location of the best 5 components is found using the SFFS algorithm, as it was explained in section 5.4.

Table 5-3 The results of the experiments related to Fig. 5.8.

<table>
<thead>
<tr>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>63</td>
<td>57.6</td>
<td>84</td>
<td>73</td>
<td>66.4</td>
<td>86</td>
<td>78</td>
<td>52.4</td>
<td>87.8</td>
</tr>
<tr>
<td>64</td>
<td>54.8</td>
<td></td>
<td>74</td>
<td>55.4</td>
<td></td>
<td>79</td>
<td>55.4</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>72.1</td>
<td></td>
<td>75</td>
<td>71.5</td>
<td></td>
<td>80</td>
<td>69.5</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>60.2</td>
<td></td>
<td>76</td>
<td>64.6</td>
<td></td>
<td>81</td>
<td>71.0</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>57.2</td>
<td></td>
<td>77</td>
<td>56.6</td>
<td></td>
<td>82</td>
<td>58.0</td>
<td></td>
</tr>
</tbody>
</table>
Figs. 5.9 and 5.10 show some combinations of the 6 and 7 components schemes that are used in our system. The recognition rates that we obtained here are 92% and 93.5%, respectively. Note that the more components overlap and the more components are used the better performance we get. Tables 5.4 and 5.5 list some of the results on experiments using 6 and 7 components of an image, respectively.

Figure 5-8 Some combinations of the five components face image of size 20×20.
Figure 5-9 Some combinations of the six components face image of size $20\times20$. 
Figure 5-10 Some combinations of the seven components face image of size 20×20.

Table 5-4 The results of the experiments related to Fig. 5.9.

<table>
<thead>
<tr>
<th>Comp.</th>
<th>Success</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success</th>
<th>Fusion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>164</td>
<td>54.6</td>
<td>92</td>
<td>170</td>
<td>52.4</td>
<td>89</td>
<td>176</td>
<td>52.4</td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>55.8</td>
<td></td>
<td>171</td>
<td>55.4</td>
<td></td>
<td>177</td>
<td>55.4</td>
<td></td>
</tr>
<tr>
<td>166</td>
<td>75.1</td>
<td></td>
<td>172</td>
<td>71.5</td>
<td></td>
<td>178</td>
<td>71.5</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>59.2</td>
<td></td>
<td>173</td>
<td>65.6</td>
<td></td>
<td>179</td>
<td>48.0</td>
<td></td>
</tr>
<tr>
<td>168</td>
<td>53.2</td>
<td></td>
<td>174</td>
<td>43.6</td>
<td></td>
<td>180</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>169</td>
<td>64.4</td>
<td></td>
<td>175</td>
<td>48.0</td>
<td></td>
<td>181</td>
<td>63.4</td>
<td></td>
</tr>
</tbody>
</table>

Table 5-5 The results of the experiments related to Fig. 5.10.

<table>
<thead>
<tr>
<th>Comp.</th>
<th>Success</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success</th>
<th>Fusion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>211</td>
<td>58.6</td>
<td>92</td>
<td>259</td>
<td>52.4</td>
<td>93.5</td>
<td>273</td>
<td>53.4</td>
<td></td>
</tr>
<tr>
<td>212</td>
<td>56.4</td>
<td></td>
<td>260</td>
<td>55.4</td>
<td></td>
<td>274</td>
<td>58.4</td>
<td></td>
</tr>
<tr>
<td>213</td>
<td>76.5</td>
<td></td>
<td>261</td>
<td>71.5</td>
<td></td>
<td>275</td>
<td>72.5</td>
<td></td>
</tr>
<tr>
<td>214</td>
<td>60.5</td>
<td></td>
<td>262</td>
<td>65.6</td>
<td></td>
<td>276</td>
<td>56.0</td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>55.2</td>
<td></td>
<td>263</td>
<td>67.6</td>
<td></td>
<td>277</td>
<td>57.0</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>66.4</td>
<td></td>
<td>264</td>
<td>59.1</td>
<td></td>
<td>278</td>
<td>66.4</td>
<td></td>
</tr>
<tr>
<td>217</td>
<td>62.6</td>
<td></td>
<td>265</td>
<td>62.3</td>
<td></td>
<td>279</td>
<td>70.4</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 5.11 shows a comparison between the recognition rates for different number of components (all of size 20×20). Note that by using 14 components, a success recognition rate of 98.5% was obtained.

![Figure 5-11 A comparison between the recognition rates for different number of components using the xm2vts database.](image)

Table 5.6 shows the amount of time consumed by 2.8 GHz – 512k RAM machine to perform the recognition process for the different number of components and for the component LDA method that was presented in [22], when applied to the xm2vts database taking 1180 images (stages one and two). A tremendous reduction in the amount of time is clearly achieved by our proposed method with a success rate of 98.5% in less than 17 minutes, for the 14 components case. Whereas the method proposed in [22] takes more than 83 minutes to achieve similar performance.

<table>
<thead>
<tr>
<th>Number of Components (all components are of size 20×20)</th>
<th>4 Opt. comp.</th>
<th>5 Opt. comp.</th>
<th>6 Opt. comp.</th>
<th>7 Opt. comp.</th>
<th>14 Opt. comp.</th>
<th>5 Comp. [22]</th>
<th>14 Comp. [22]</th>
</tr>
</thead>
<tbody>
<tr>
<td>time (min)</td>
<td>4.9</td>
<td>6.7</td>
<td>7.8</td>
<td>8.6</td>
<td>16.9</td>
<td>72.2</td>
<td>83.4</td>
</tr>
<tr>
<td>Success Rate %</td>
<td>81.4</td>
<td>87.8</td>
<td>92</td>
<td>93.5</td>
<td>98.5</td>
<td>Not mentioned</td>
<td>98.5</td>
</tr>
</tbody>
</table>
To prove the efficiency of our proposed method, we present the performance statistics and report them as cumulative match scores (CMS). In this case, identification is regarded as correct if the true object is in the top Rank $n$ matches (as it has been explained in chapter 4). The results for the xm2vts (for stage one taking 4 images from the first shot) and ORL databases are shown in Figs. 5.12 and 5.13, respectively. It is worth mentioning that these curves were also calculated for MPEG-7, Holistic LDA, Eigenphases, and PCA techniques as shown in the figures.

![Figure 5-12 Comparison of the CMS curves of various methods for the xm2vts database (First shot with 1180 images).](image-url)
Figure 5-13 Comparison of the CMS curves of various methods for the ORL database.
In addition, the receiver operating characteristic (ROC) curve is used to assess the performance of the proposed method and compared to different other techniques. The ROC curves for the xm2vts (for stage one taking 4 images from the first shot) and ORL databases are shown in Figs. 5.14 and 5.15, respectively.

![ROC curves comparison](image)

Figure 5-14 Comparison of the ROC curves of various methods for the xm2vts database (First shot with 1180 images).
Figure 5-15 Comparison of the ROC curves of various methods for the ORL database.
To show that the performance of the system is stable, Figs. 5.16 and 5.17 show the CMS and the ROC curves for the different methods for the second stage of xm2vts experiments, where we have used the other 4 images that were taken from the second shot. Thus, we have another 1180 images for the 295 different individuals. Comparing the first and the second stage performances (see Figs. 5.12, 5.14, 5.16, and 5.17), we see that both stages give very high and significant results.

Figure 5-16 Comparison of the CMS curves of various methods for the xm2vts database (Second shot with 1180 images).
Figure 5-17 Comparison of the ROC curves of various methods for the xm2vts database (Second shot with 1180 images).
Further, Figs. 5.18 and 5.19 illustrate the CMS and the ROC curves for the different methods for the third stage of xm2vts experiments, where the whole database is used (2360 images for 295 different individuals). From Figs. 5.12, 5.14, 5.16, 5.17, 5.18, and 5.19, we can comfortably state that the performance of the system is robust for the three stages, with very high and significant results. It is important to mention that to compare the results of stage three to stage one and two, we had to increase the number of eigenvectors that are needed to represent the face space. In the case of stage one and stage two, we used 10 eigenvectors to attain high performance. Whereas in the case of the third stage we had to use 30 eigenvectors corresponding to the largest 30 eigenvalues in order to represent the system effectively. To further emphasize on the issue of the eigenvectors, an experiment is conducted where the number of eigenvector is increased and the corresponding performance is recorded for the third stage.

![Figure 5-18 Comparison of the CMS curves of various methods for the xm2vts database (2360 images).](image-url)
Figure 5-19 Comparison of the ROC curves of various methods for the xm2vts database (2360 images).
Fig. 5.20 shows the results of increasing the number of the eigenvectors on the performance of the system. The curve shown is a sample for a single component of size 20×20 when all the 2360 images from the xm2vts database are used. We see that 10 eigenvectors give poor result. The performance improves when the number of eigenvectors encountered is increased until it reaches 50. After that the performance of the system gets saturated even if we increase the number of the eigenvectors to 150. Finally, when we add more eigenvectors to our system, the performance degrades and this can be seen from the curve when the number of eigenvectors goes beyond 150.

![Graph showing the relation between the number of eigenvectors and the performance of the system.](image)

**Figure 5-20** Relation between the number of eigenvectors encountered and the performance of the system.
Table 5.7 shows the amount of time consumed by 2.8 GHz – 512k RAM machine to perform the recognition process for the different number of components and for the method presented in [22], when applied to the xm2vts database using 2360 images. As expected, a system with 2360 images will take more time for recognition than a system with 1180 images.

However, our proposed method performs the recognition process much faster than the method proposed in [22]. Our method gives a success rate of 98.5% in only ~27 minute, for the 14 components case; whereas the method proposed in [22] takes more than 134 minutes to achieve the corresponding high rate. In other word, our technique performs 5 times faster than the one proposed in [22].

Note that, although the 30×30 four-components method achieves 99.5% success rate, it takes almost 49 minutes to accomplish its task. Whereas the 20×20 14-components method achieves 98.5% success rate in only 27 minutes, that is it needs approximately half the time that the 30×30 four-components needs. Another observation that can be made from the table (columns 2, 3, and 4) is that the success rate decreases by decreasing the component size (with the number of components kept unchanged). Finally, fixing the component size (columns 4, 5, 6, 7, and 8) and increasing the number of components will lead to an increase in the success rate.

Table 5-7 The time consumed (in minutes) and the success recognition rate for the different number of components and the component LDA method described in [22] using 2360 images of the xm2vts database.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Success Rate %</td>
<td>99.5</td>
<td>99.2</td>
<td>80</td>
<td>86.9</td>
<td>91.1</td>
<td>92.4</td>
<td>98.5</td>
<td>Not mentioned</td>
<td>98.5</td>
</tr>
</tbody>
</table>
To take a closer look at the complexity analysis of the component parts of the new algorithm, we provide a plot of computation cost (time) for the LDA process for a single component as a function of window size. Fig. 5.21 shows this time for a single component from window size of 56×46 (full image size) down to size of 10×10 window. As seen from the plot, the processing time drops dramatically as the window size decreases to 35×35. Then, it starts to decrease at a slower rate until the size of the image reaches 20×20. An initial remark tells us that it is not feasible to think of components of size larger than 30×30, because the processing time for a single component of such a size will take at least 13 minutes. Since we need to have at least four components in order to have an attractive performance, the total processing time will be around 50 minutes, as it is clearly demonstrated in Table 5.7.

![Figure 5-21 Computation time for the LDA process for a single component as a function of window size.](image)

A closer look at the relation between the component size and the corresponding processing time is given in Fig. 5.22, where the plot shows the results for components starting from 30×30 size. Now, note that the processing of a single component of size 25×25 takes around 7.5 minutes, whereas the processing time for a single component of size 20×20 takes approximately 1.8 minutes. So to achieve the performance results shown in Table 5.7, one needs to process four components each of size 25×25 which takes at least 30 minutes (actually the system takes more than this time when it process all the four components at the same time), whereas it takes around 25 minutes to process 14 components each of size 20×20. This observation leads us to choose
component size 20×20 when we seek the fastest processing. Further, if we take the performance factor in consideration, we see that going below the size of 20×20 will not give a good performance. This will be shown in the next subsection when we present the performance results of component size 15×15. In brief, taking both factors together (the best performance and the fast processing time), we reach a reasonable conclusion that the component size 20×20 is the best that can represent our system.

![Figure 5-22 A closer look at the computation time for the LDA process for a single component as a function of window size starting from 30×30.](image)

In addition, one of the considerations affecting our design methodology is the computational complexity. In describing the computational complexity of an algorithm, we are generally interested in the number of basic mathematical operations, such as additions and multiplications it requires, beside the time and memory needed on a computer. To illustrate this computational complexity concept for the component-based LDA method, Table 5.8 shows the number of multiplications and additions required to build the covariance matrix and the complexity of finding the Fisherfaces. Note that the computational complexity of calculating the Fisherfaces is of the order of $O(Ld^3)$, where $L$ is the number of components used to represent the recognition system and $d$ is the component size [23, 165]. From Table 5.8, we can notice the following:
a) The computational complexity of the full image size (56×46) is very high
compared to other components schemes.

b) As the component size reduces, the corresponding complexity reduces
dramatically.

c) For a certain image size (for example 20×20), as the number of components
encountered in the processing increases, the total corresponding computational
complexity also increases.

d) The complexity of processing four components of size 25×25 is greater than
the complexity of processing 14 components of size 20×20.

Finally, the above results and remarks are in agreement with our previous results
presented in Table 5.7 which demonstrates the time consumed for the processing of
the component-based LDA system.

Table 5-8 Comparison of the computational complexity between the different components
schemes. Note: \( d = \) dimensionality of the image = \( m \times n \), and \( M = \) the number of images used in
training= 1180.

<table>
<thead>
<tr>
<th>Component size = m×n</th>
<th>Number of multiplications needed for a single component</th>
<th>Number of additions needed for a single component</th>
<th>Number of components used</th>
<th>Computational complexity to find Fisherfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>( n )</td>
<td>( n \times m \times M^2 )</td>
<td>( (n \times m - 1) \times M^2 )</td>
<td>( L )</td>
</tr>
<tr>
<td>Holistic LDA</td>
<td>56</td>
<td>46</td>
<td>3,586,822,400</td>
<td>3,585,430,000</td>
</tr>
<tr>
<td>4 Components</td>
<td>30</td>
<td>30</td>
<td>1,253,160,000</td>
<td>1,251,767,600</td>
</tr>
<tr>
<td>4 Component</td>
<td>25</td>
<td>25</td>
<td>870,250,000</td>
<td>868,857,600</td>
</tr>
<tr>
<td>4 Component</td>
<td>20</td>
<td>20</td>
<td>556,960,000</td>
<td>555,567,600</td>
</tr>
<tr>
<td>5 Component</td>
<td>20</td>
<td>20</td>
<td>556,960,000</td>
<td>555,567,600</td>
</tr>
<tr>
<td>6 Component</td>
<td>20</td>
<td>20</td>
<td>556,960,000</td>
<td>555,567,600</td>
</tr>
<tr>
<td>7 Component</td>
<td>20</td>
<td>20</td>
<td>556,960,000</td>
<td>555,567,600</td>
</tr>
<tr>
<td>14 Component</td>
<td>20</td>
<td>20</td>
<td>556,960,000</td>
<td>555,567,600</td>
</tr>
</tbody>
</table>
5.5.2.1 The Components Size and Location

The important issue of the size of the selected components in a face image is studied in order to reach some conclusions. This section presents some experiments that were carried out on the xm2vts database.

For the sake of comparison, we have conducted a few experiments taking four, five, and six components from the face image but with component size of 15×15. Samples of the components are shown in Figs. 5.23, 5.24, and 5.25. Samples of the results of these experiments corresponding to those figures are listed in Tables 5.9, 5.10, and 5.11, respectively. From these tables, we can conclude that increasing the number of components while making their size smaller does not improve the recognition rate.

![Figure 5.23 Four components face image of size 15x15.](image)

Table 5.9 The results of the experiments related to Fig. 5.23, the 4 components 15x15 patches.

<table>
<thead>
<tr>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>31.5</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>30.2</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>22.9</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>29.3</td>
<td>33.22 %</td>
</tr>
</tbody>
</table>
Figure 5-24 Different combinations of five components face image each of size 15×15.

Table 5-10 The results of the experiments related to Fig. 5.24, the 5 components 15×15 patches.

<table>
<thead>
<tr>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success %</th>
<th>Fusion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>88</td>
<td>26.6</td>
<td>47.1</td>
<td>93</td>
<td>29.7</td>
<td></td>
<td>98</td>
<td>29.3</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>28.0</td>
<td></td>
<td>94</td>
<td>28.3</td>
<td></td>
<td>99</td>
<td>27.1</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>40.2</td>
<td></td>
<td>95</td>
<td>16.3</td>
<td></td>
<td>100</td>
<td>38.5</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>29.8</td>
<td></td>
<td>96</td>
<td>38.6</td>
<td></td>
<td>101</td>
<td>38.6</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>28.0</td>
<td></td>
<td>97</td>
<td>16.4</td>
<td></td>
<td>102</td>
<td>28.6</td>
<td></td>
</tr>
</tbody>
</table>

On the other hand, the experiments have demonstrated that increasing the component size has much better effect on the performance than increasing the number of the components. This was demonstrated in Table 5.1 for the case of 30×30 and 25×25 and Tables 5.2-5.5 for the case of 20×20 component size. However, this performance will be obtained at the expense of speed and time efficiency. The execution time for the 4 components LDA of size 20×20 is higher than the corresponding time for the 5 components LDA of size 15×15.
Further, a closer look at the results obtained from the 6 component LDA (Table 5.11) that corresponds to the non overlapped components in Fig. 5.25 (top-left case, component # 194), and comparing these results to the ones with overlapped components shown in the same figure (lower-right one, component # 227), we can draw a very important conclusion: overlapping of components improves the recognition rate. The reason is that overlapping preserves the relation between adjacent components, which in turn enhances the recognition process.

Table 5-11 The results of the experiments related to Fig. 5.25, the 6 components 15×15 patches.

<table>
<thead>
<tr>
<th>Comp.</th>
<th>Success</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success</th>
<th>Fusion Rate</th>
<th>Comp.</th>
<th>Success</th>
<th>Fusion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>194</td>
<td>30.3</td>
<td>44.9</td>
<td>218</td>
<td>37.8</td>
<td>48.8</td>
<td>224</td>
<td>28.8</td>
<td>49.7</td>
</tr>
<tr>
<td>195</td>
<td>28.6</td>
<td></td>
<td>219</td>
<td>33.2</td>
<td></td>
<td>225</td>
<td>28.1</td>
<td></td>
</tr>
<tr>
<td>196</td>
<td>26.3</td>
<td></td>
<td>220</td>
<td>33.6</td>
<td></td>
<td>226</td>
<td>24.6</td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>24.4</td>
<td></td>
<td>221</td>
<td>22.7</td>
<td></td>
<td>227</td>
<td>37.8</td>
<td></td>
</tr>
<tr>
<td>198</td>
<td>23.4</td>
<td></td>
<td>222</td>
<td>19.2</td>
<td></td>
<td>228</td>
<td>33.2</td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>26.1</td>
<td></td>
<td>223</td>
<td>26.8</td>
<td></td>
<td>229</td>
<td>26.8</td>
<td></td>
</tr>
</tbody>
</table>
5.6 Conclusion

In this chapter, a new approach for finding the optimal component-based LDA face representation was presented. The approach relates the size, the number, and the location of the components in order to describe the face image efficiently, and consequently achieve the best recognition rate within the minimum possible amount of time. We have experimentally demonstrated the trade-off between the component size and the number of such components. The efficient and well known SFFS feature selection method has been adopted as part of our proposed technique. In addition, thorough experiments have been conducted in order to decide the suitable component size which can give high performance and at an acceptable processing time. It has been found that component size of less than 20×20 would not give good recognition rate. Moreover, we experimentally demonstrated that increasing the component size reflects directly on the processing time. In other words, a trade-off between the component size and the time to process them must be reached. A large number of experiments involving components with different sizes, numbers, and locations have been conducted. Moreover, our proposed method was compared to other published well known methods and it was shown that we have achieved much better performance.

Finally, the proposed technique and methodology have important practical applications as it helps to find the best way to extract facial features and represent the face image. Some of these applications are face recognition under occlusion problem, limited memory, fast speed devices, and other important constraints that will be discussed in the next chapter.
Chapter 6
Discussions and Applications

As we have shown throughout the thesis and basically from the experiments that we have presented in chapters 4 and 5, the face recognition field is very challenging. The broad spectrum of applications of this field of science make it attractive to researchers from different disciplines. In this chapter, we try to comment on some of the results that we obtained, and present some of the application areas that we think our proposed system is suitable for implementation.

First of all, we tried to propose a full range of solutions for face recognition. The strong motivation behind our work is the rapid development in electronics and communication engineering as well as the hardware technology. The applications that were difficult to implement in the past are now coming very close to being employed.

There are three major issues for face recognition which we addressed in this research: how can we get good performance? how fast can we get this performance? and how much memory is required to get this performance?

Apparently obtaining the best answers for the above three questions is our desired main goal. However, this objective may not be practically achievable and consequently the issue of optimality is usually raised, where a trade-off between these criteria is imposed. Thus, we can always accept a certain performance for the benefit of another criterion. For instance, in chapter 4 we introduced binarized eigenphase technique, where we focused on obtaining a high level of performance while maintaining the high speed of the system at the same time. But these two objectives cannot be achieved efficiently without having a system with a huge memory during processing and signal transmitting. Therefore, knowing that wireless communication systems have a limited memory capacity will impose on us the requirement of fast processing rates at the expense of less performance.

It is worth mentioning that for the technique proposed in chapter 4 we tried in some early experiments to use the LDA instead of the PCA. However, the results were not encouraging at all and the system did not perform well. The LDA is well
known to be effective for spatial or image domain features computed from pixel intensity values, but, to the best of our knowledge, its effectiveness is not proved yet in the frequency domain.

In chapter 5, we tried to overcome the relatively low recognition rate of the system. We managed to develop a very effective system that outperforms the binarized eigenphase one of chapter 4, and we achieved a very high performance. However, when compared to the binarized eigenphase system, an improved recognition performance was achieved at the expense of a large amount of computation time that is needed. Although large, this consumed time is still much less than the time consumed by other known techniques, as it has been shown and demonstrated in chapter 5.

To summarize, the two proposed methods in chapters 4 and 5 complement each other. Even though the binarized eigenphase method is an extended and developed version of the MPEG-7 technique, the proposed optimal component-based LDA technique can be a good candidate for the MPEG-7 face descriptor for multimedia content description.

In the next subsections, we present some potential applications for the proposed methods for face recognition.

A. Limited Memory Applications

Recently, the wearable computing field is rapidly expanding and attracting more and more people interested in the field. Computers and computing devices, cameras, microphones, sensors and many other devices are subject to being built into a person’s clothes [119, 124]. These wearable devices can be adapted to a specific user and be more closely and actively involved in the user’s activities. Consequently, we can expect to see a rapidly growing interest in image interpretation in this widely unexplored area.

Therefore, face recognition could be an integral part of wearable systems like memory aids and context-aware systems. Developers are expected to integrate many future recognition systems with clothing and accessories. For instance, if you build a camera into an eyeglass, then face recognition software can help in recognizing the name of the person by whispering in the ear.
Obviously the above mentioned applications cannot offer a large memory, and consequently we should look for techniques for face recognition that utilizes compact small memories. Our proposed method in chapter 4 can be one alternative solution since the binarization means that we do not need more than one bit to represent a pixel. This is in contrast to 5 bits, which are needed by MPEG-7 proposals for face description.

**B. Small and Portable Devices**

Despite the significant advances in face recognition technology, it has yet to achieve the levels of accuracy required for many commercial and industrial applications. With the increased worldwide spread in portable devices and with the need of advanced real-time face recognition systems, the urgent need for more cost effective and low power consumption of such systems has increased. Typical face recognition algorithms may not be executable with current memory-constrained devices. However, the fast advancement of the technology of the wireless communication devices on one hand, and the huge progress in the area of small memory hardware technology on the other hand, both have motivated researchers to explore new techniques and systems for face recognition that might be launched on small and portable devices. These portable devices, such as mobile phones, have become an integrated part of our lives where they do more than just communicating between people. Nowadays most of the existing mobile phones are integrated with a camera which makes it very convenient to implement a face recognition algorithm on such portable devices. Mobile phone based face recognition systems can be used in a variety of applications including law enforcement officers, who can use mobile phones to identify suspects against a criminal watch list, for example. Further, mobile phones can be used as a way to delegate authority and access control to vital places.

**C. Mobile Devices Applications**

The transmission and receiving of digital information from and to wireless devices are very exciting recent developments. Mobile devices such as cellular phones, Personal Digital Assistant (PDA) or laptops which are equipped with web browsers and mail software have become popular tools to access the Internet for both personal and business use. Moreover, owing to the arrival of the IMT-2000 communication service,
the wireless network bandwidth is expected to rapidly widen. The first implementation stage of this service will offer broadband packet switched wireless channels of up to 384 kbps. This will enable the delivery of high quality audiovisual (AV) content to mobile users in the near future. Thus, mobiles devices become part of our lives, and with a fast growing real-time digital information processing, we will not be surprised to see face identification and verification applications to be implemented soon on these mobile devices. The proposed techniques of this thesis would be very good candidates for practical implementations in mobile devices.

D. Wireless Surveillance Systems

The wireless surveillance systems cover many applications ranging from traffic problems to security ones. An average citizen is exposed to video surveillance in many day-to-day activities such as banking, shopping, gas purchasing, driving through major road intersections, and entering public and private buildings, to just name a few. There is a great inherent appeal in the idea that future terrorists' attacks could be prevented by high-tech video surveillance.

The traditional scene of a policeman that carries along a photo of a wanted person and visually scans crowds to identify the person will be soon replaced by new portable automated face recognition systems. The proposed face recognition techniques of chapters 4 and 5 could represent possible solutions. The use of face recognition technology makes it feasible to do this type of scanning where it is difficult in practical terms, using human observers.

Another major difficulty that wireless systems may face is the tremendous amount of data which needs to be uploaded and downloaded from and to the portable devices. This requirement imposes a strong constraint to the portable devices in terms of their memory capacity.

The methods proposed in this thesis may find a good and interesting platform in these types of systems and applications: where we can identify and verify some criminals or some monitored persons.
E. MPEG-7 Face Descriptors

Recently, many face descriptors for MPEG-7 have been proposed for face image retrieval in video streams. An important characteristic of these descriptors is their compactness [16, 17, 18, 124]. The binarized eigenphase technique which is demonstrated in chapter 4 represents a new development for face description for MPEG-7. On the other hand, component-based LDA presented in chapter 5, can also be another method for face representation for MPEG-7 face image feature descriptor. This is mainly because the component-based LDA provides solutions for the problems of face retrieval and person identification using a vector of small size and of low dimension.
Chapter 7
Conclusions and Future Work

7.1 Conclusions

Face recognition has numerous commercial and law enforcement applications. Although humans have the ability to recognize faces in cluttered scene with relative ease, machine recognition is a much more difficult dilemma. During the last few years machine recognition of faces has emerged as an active research area spanning different disciplines such as image processing, pattern recognition, computer vision, computer graphics, and neural networks. In chapter 2, we have presented a survey for the most famous and the state-of-the-art techniques for face recognition that found great applications in today's life. In chapter 3, we have presented the mathematical emphasis for the two important classification techniques (PCA and LDA). The heavy potential use of the face as an identification and verification tool makes it one of the most important and powerful biometrics measures. Although many face recognition techniques have been proposed and have shown significant promises, robust face recognition is still a challenging task.

With the rapid advancement in the field of digital communication and wireless transmission, along with widespread of using mobile phones and small devices, there is a great potential for important and vital applications of face recognition systems. The constraints of the high speed and the low memory requirements for such systems challenge the researchers to come up with new techniques and/or improve the existing ones, in order to cope with such hard constraints and difficult conditions.

Moreover, multimedia data search and retrieval has become a very active research field because of the increasing amount of audiovisual (AV) data available to the system. The growing difficulties to search, filter, or manage data and the needs for interoperability between devices have also been recognized. Several standardization activities have been launched and MPEG-7 is being one of them that standardises the description of multimedia content to support wide range of applications.
One of the advantages of MPEG-7 is that it allows the extension of the core set of description tools in a standard way. The standard can never contain all the structures needed to address all the applications in the entire different domain, and thus it is possible to extend the standard in a way that guarantees as much interoperability as possible.

In this thesis, novel methods are proposed for efficient face recognition that can be implemented in systems that have limited memory capabilities and have low speed processors. As explained in chapter 4, the new technique exploits the characteristics and advantages of the MPEG-7 FFD vectors, frequency domain, binarization process, and the PCA. Moreover, the new technique enhances the performance of the face recognition by providing a much better recognition rate compared to other known techniques when applied to two independent face images databases. In addition, our proposed technique offers another important advantage by providing a low storage capability. Despite the remarkable advantages that we have obtained, we believe that this track of research needs further exploration to cope with the problems of image illumination and poor resolution.

Furthermore, we have thoroughly studied in chapter 5 the component-based LDA for face recognition. We found the optimal component-based LDA face representation through finding the optimal size, number, and location of components needed to describe the face image efficiently, within the minimum possible amount of time. Note that the problem of feature selection in the context of computer vision involves the following issues: large number of features, many irrelevant features, many redundant features, and noisy data. Feature selection is a searching problem that detects an optimal feature subset based on the selected measure. Many algorithms have been proposed for feature selection. In our study of component-based LDA; the SFFS procedure as a feature selection method was used to find the optimal components. The method achieved very high rates of recognition and verification. Also, we have suggested the component-based LDA as a possible solution for the occlusion problem as well as for systems with limited memory and speed.

Chapter 6 of the thesis presented some of the applications where the new proposed systems can find a place for implementation. Among these areas: limited memory application, small and portable devices, mobile devices, wireless surveillance systems, and MPEG-7 face descriptor.
It is worth mentioning that the work proposed in chapter 4 was published and presented in the following conferences:

- International Conference on Visualization, Imaging, and Image Processing, Spain, 2006
- International Workshop on Imaging Systems and Techniques, Italy, 2006
- International Conference on Computer Graphics and Imaging, Austria, 2007
- International Conference on Computer Vision Systems, Germany, 2007

where only 1180 images from the xm2vts database were used at that time. However, the work presented in that chapter has utilized the whole xm2vts database (2360 images), beside the ORL one. Also, the work was submitted to a journal where it is currently under review. Moreover, the component-based LDA presented in chapter 5 is submitted for publication, as well.

### 7.2 Future Work

It might seem that the face recognition has reached a mature level and acceptable performance, but actually there are many associated problems that need to be solved.

- The performance of the proposed system in chapter 4 can be challenged when it is used with new databases of image with different illuminations and pose variations. For instance, outdoor lighting causes a more complicated face appearance variation than indoor and controlled lighting, since it involves multiple light sources from sky and the reflectance from other objects, and the human may also be placed in the shadows of other objects. Future works can be pursued to provide solutions for correcting lighting variations for face recognition. A direct consequence result of solving this problem, along with our proposed technique, is a practical implementation in the field of surveillance systems. One possible solution that addresses lighting variations might be to apply Logarithm transformation [46]. Logarithm transformation can non-linearly map the pixel intensities; as a result the shadowed regions will be intensity enhanced producing a final image with more distinctions in the low-illumination regions.
• Also, as we have seen and discussed in chapter 4, we have utilized the median filter for the binarization step in our proposed technique. We can explore other filters and study the possibility of achieving better results.

• A possible future direction for this research is the use of IPCA and/or the ICA schemes (which are briefly discussed in chapter 2), in the proposed technique of chapter 4. An attractive feature of the IPCA-ICA is that a small memory is needed to store the whole data matrix that represents the incoming images, since the next incoming images will be stored over the previous images and vectors. This fits very well with the requirements of our proposed systems that we have discussed earlier.

• Further the super-resolution algorithm [47, 51] can be investigated in our system where it might be embedded into our face recognition system to solve the problem of poor resolution. As we mentioned in chapter 2, super resolution is not performed in the pixel domain, but is instead performed in a reduced dimensional domain. By this the computation complexity is reduced. In other words, the face observations are first projected to the face space, and then the super-resolution reconstruction is performed in the low-dimensional face subspace instead of the spatial domain.

• With respect to the proposed component-based LDA technique, a further study can be carried out to investigate the performance of the system under more severe conditions where the face images are under a varied pose. Employing Gabor filtering [83, 84] to each of the components before applying the LDA technique might be a direction that worth more investigation.

• In addition, it will be interesting to fuse different modalities into one single system instead of relying on a single modality. 3D systems can play a good role in this, where we may combine 2D and 3D systems to provide a better system.

• A further study to explore the frequency domain characteristics of the LDA system can also be investigated.

• Moreover, facial asymmetry is a challenging and interesting topic and it is one possible direction for our future work. As it was explained in chapter 2, Facial asymmetry can be caused either by external factors such as expression
changes, viewing orientation and lighting direction, or by internal factors such as growth, injury, and age related changes. Frequency domain characteristics can be useful in this analysis [108].
Appendix
MPEG-7: Multimedia Content Description Interface

A. Introduction

Multimedia search and retrieval has become a very active research field because of the increasing amount of audiovisual (AV) data that needs to be handled. The growing difficulty to search, filter, or manage such data imposes practical challenges. Furthermore, many new practical applications such as large-scale multimedia search engines on the Web, media asset management systems in corporations, AV broadcast servers, and personal media servers for consumers are about to be widely available. This framework has led to the need of developing of efficient processing tools that are able to create the description of AV material or to support the identification or retrieval of AV documents. Beside the research activity on processing tools, the need for interoperability between devices has also been recognized and several standardization activities have been launched. MPEG-7 [16, 124], standardises the description of multimedia content supporting a wide range of applications. Standardization activities do not focus so much on processing tools but concentrate on the selection of features that have to be described, and on the way they are structured and instantiated with a common language.

Using digital cameras with personal computers and the Internet, virtually every individual in the world is a potential content producer, capable of creating content that can be easily distributed and published, and to be available on-line.

It is necessary to automatically and objectively describe, index and annotate multimedia information, notably audiovisual data, using tools that automatically extract audiovisual features from the content to substitute or complement manual, text-based description. These automatically extracted features will have three advantages:

1. they will be automatically generated,
2. they can be more objective and domain-independent, and
3. they can be native to audiovisual content.

The MPEG-7 project has the objective to specify a standard way of describing various types of multimedia information such as elementary pieces and complete work, irrespective of their representation format and storage medium. These descriptions are both textual (annotations, names, etc.) and nontextual (statistical features, camera parameters, etc.). Like the other members of the MPEG family, MPEG-7 defines a standard representation of multimedia information satisfying a set of well-defined requirements. But MPEG-7 is quite a different standard than its predecessors: MPEG-1, MPEG-2 and MPEG-4 which represent the content itself—'the bits'. The MPEG-7 represents information about the content—'the bits about the bits'. While the first reproduce the content, the latter describes the content. The requirements for these two purposes are very different. This standard allows content and its descriptions to be exchanged across different systems and can set an environment where tools from different providers can work together, creating an infrastructure for transparent management of multimedia content. One of the advantages of MPEG-7 is that it allows the extension of the core set of description tools in a standard way. The standard can never contain all the structures needed to address all the applications in the entire different domain, and thus it is possible to extend the standard in a way that guarantees as much interoperability as possible.

The MPEG-7 standard will be extended with additional tools to address more requirements and provide more functionality, where the technologies best addressing the requirements of certain application are selected to replace or complement the old ones.

B. MPEG-7 Types of Tools
It is clear that a number of different tools are needed to achieve the standard's objectives. These tools are descriptors (the elements), description schemes (the structures), a Description Definition Language (DDL) (for extending the predefined set of tools) and a number of Systems tools [143].

Descriptors: a descriptor is a representation of a feature, where a feature is a distinctive characteristic of the data (audiovisual information that will be described)
that signifies something to somebody. It defines the syntax and the semantics of the feature representation.

**Description Schemes:** specifies the structure and semantics of the relationships between its components, which may be both descriptors and description schemes. As an example for this: a movie (*Description Scheme*) is structured as scenes and shots, including some textual *descriptors* at the scene level and colour, motion and audio amplitude *descriptors* at the shot level.

**Description Definition language (DDL):** is the language that allows the creation of new description schemes and (possibly) descriptors. DDL allows the extension and modification of existing description schemes.

**System Tools:** related to the binarization, synchronization, transport and storage of descriptions, as well as the management and protection of intellectual property.

There are some requirements for the MPEG-7 tools that they should adhere to. Cross-modality is one of the requirements for MPEG-7 descriptors, where they support audio, visual or other descriptors which allow queries based on visual descriptions to audio data and vice versa. For the description scheme, the different schemes shall express the relationships between descriptors to allow the use of the descriptors in more than one description scheme. Moreover, the DDL will allow descriptors and description schemes to be created and existing one to be modified or extended. In addition, it provides unique identification of descriptors and description schemes. Also, DDL supports a rich model for links and references between one or more descriptions. DDL allow descriptors and description schemes to be readable by humans. MPEG-7 extensibility allows users to expand MPEG-7 to suit their own specific needs and the standard to keep evolving, integrating novel description tools.

### i. Systems Architecture

The concept of ‘Systems’ in MPEG has evolved dramatically since the development of the MPEG-1 and MPEG-2 standards. In the past, ‘Systems’ referred only to overall architecture, multiplexing and synchronization. In MPEG-4, in addition to these
issues, the Systems part encompasses interactive scene description, content
description and programmability.

MPEG-7 brings new challenges to the systems expertise such as languages for
description representation, binary representation of descriptions and delivery of
descriptions either separate or jointly with the audiovisual content. The combination
of the new possibilities of describing audiovisual content offered by MPEG-7
Systems and the efficient description tools provided by the Visual, Audio and
Multimedia Description Schemes parts of the standard, assure to be the groundwork
of a new way of thinking about audiovisual information. Indeed, without content
description, audiovisual data is mainly an opaque series of bits. Only the decoding of
these bits give some information about what the data is about and what the user can
do with it. The decoding process involves, in general, complex and demanding
operations, and requires high bandwidth in networked environments.

With the use of MPEG-7 descriptors, and description schemes, MPEG-7 provides
a way to obtain information about audiovisual data without the need of performing the
actual decoding of these data. The MPEG-7 Systems specification completes the
picture by linking MPEG-7 description with the audiovisual content and providing an
efficient binary representation of the description in the best MPEG tradition.

ii. Description Definition Language

The DDL provides the foundations for the MPEG-7 standard. It provides the language
for defining the structure and content of MPEG-7 documents. The DDL is not a
modelling language such as Unified Modelling Language (UML) but a schema
language to represent the results of modelling audiovisual data, (i.e. descriptors and
description schemes) as a set of syntactic, structural and value constraints to which
valid MPEG-7 descriptors, description schemes, and descriptions must conform. It
also provides the syntactic rules, by which users can combine, extend and refine
existing description schemes and descriptors to create application-specific description
definitions or schemas. The purpose of a schema is to define a class of XML
(Extensible Markup Language) documents. MPEG-7 instances are XML documents
that conform to a particular MPEG-7 schema (expressed in the DDL) and that
describe audiovisual content.
The DDL is capable of expressing structural, inheritance, spatial, temporal, spatiotemporal and conceptual relationships between the elements within a description scheme and between description schemes. It provides a rich model for links and references between one or more descriptions and the data that it describes. It is the platform and application independent, machine-readable. XML schema has been chosen as the basis for the DDL, because of its widespread adoption as a schema language for constraining the structure and content of XML documents, its ability to satisfy the MPEG-7 DDL requirements and the ready availability of XML schema tools.

1. Binary Format

There are two possible ways to transmit MPEG-7 descriptions: either in Textual Format (TeM) or in Binary Format (BiM). While both formats provide similar functionality in general, the BiM provides an additional feature: compression of the verbose Extensible Markup Language (XML) representation used by TeM. In addition, the compressed binary format is designed in a way that allows fast searching and filtering on binary level, without decompressing the complete description stream beforehand. This feature is very important for small, mobile, low-power devices with restricted CPU and memory capabilities.

iii. Description Schemes

The MPEG-7 description schemes expand on the MPEG-7 descriptors by combining individual descriptors and other description schemes within more complex structures and by defining the relationships between the constituent descriptors and description schemes. In MPEG-7, the description schemes are categorized as related specifically to the audio or visual domain, or related generically to the description of multimedia. For example, typically, the generic description schemes correspond to permanent metadata related to the creation, production, usage and management of multimedia as well as to describing the content directly at a number of levels including signal structure, features, models and semantics. Typically, the Multimedia Description Schemes refer to all kinds of media consisting of audio, visual and textual data, whereas the domain-specific descriptors, such as those for colour, texture, shape, melody and so forth, refer specifically to the audio or visual domain.
iv. Descriptors

The goal of the MPEG-7 standard is to allow interoperable searching, indexing, filtering and access of multimedia content by enabling interoperability among devices that deal with multimedia content description. The MPEG-7 descriptors are designed to describe individual features of multimedia content. The description schemes provide complex description by integrating together multiple descriptors and description schemes. The MPEG-7 DDL provides a language for defining the description schemes and descriptors. The DDL also allows the extension and modification of the MPEG-7 standardized description schemes and descriptors.

The MPEG-7 descriptors are designed for describing the following types of information: low-level audiovisual features such as colour, texture, motion, audio energy and so forth; high-level features of semantics objects, events and abstract concepts; content management processes; and information about the storage media and so forth.

1. Visual Descriptors

The main objective of the MPEG-7 visual standard is to provide standardized descriptions of streamed or stored images or video-standardized header bits (visual low-level descriptors) that help users or applications to identify, categorize or filter images or video [144, 145]. These low-level descriptors can be used to compare, filter, or browse images or video purely on the basis of non-textual visual descriptions of the content – or in combination with common text – based queries.

Selected application examples include digital libraries (image and video catalogues), broadcast media selection (TV channels) and multimedia editing (personalized electronic news service, media authoring). Among this diversity of possible applications the MPEG-7 visual feature descriptors allow users or agents to perform the following tasks taken as examples:

- **Graphics**: Draw a few lines on a screen and get in return a set of images containing similar graphics or logos.

- **Images**: Define objects, including colour patches or textures and get in return examples among which you select the ones of interest.
- **Video**: On a given set of video objects, describe object movements, camera motion or relations between objects and get in return a list of videos with similar or dissimilar temporal and spatial relations.

- **Video activity**: On a given video content, describe actions and get a list of videos in which similar actions happen.
References


125


