TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

Pulchowk Campus

Department of Electronics and Computer Engineering

Thesis No: IC(2063MSI609)

"Comparative Analysis of Face Recognition Methods"

by

MANISH SHRESTHA

(Roll No. 063/MSI/609)

A thesis submitted to the department of Electronics and Computer Engineering in partial fulfillment of the requirements for the Degree Master of Science in Information and Communication Engineering

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

LALITPUR, NEPAL

SEPTERMBER, 2010

Comparative Analysis of Face Recognition Methods

by

Manish Shrestha

Roll No: 063/MSI/609

Thesis Supervisor

Asst. Prof. Daya Sagar Baral

Institute Of Engineering

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Information and Communication Engineering

Department of Electronics and Computer Engineering

Institute of Engineering, Pulchowk Campus

Tribhuvan University

Lalitpur, Nepal

September 2010

Recommendation

The undersigned certify that they have read and recommended to the Institute of Engineering for acceptance, a thesis entitled **"Comparative Analysis of Face Recognition Methods**", submitted by **Manish Shrestha** in partial fulfillment of the requirement for the award of the degree of "**Master of Science in Information and Communication Engineering**".

……………………………………

Mr. Daya Sagar Baral, Asst. Prof.

Institute Of Engineering

Supervisor

……………………………………

Mrs. Pratima Pradhan

Nepal Telecom

External Examiner

Departmental Acceptance

The thesis entitled **"Comparative Analysis of Face Recognition Methods",** submitted by **Manish Shrestha** in partial fulfillment of the requirement for the award of the degree of "**Master of Science in Information and Communication Engineering**" has been accepted as a bonafide record of work independently carried out by him in the department.

Prof. Dr Shashi Dhar Ram Joshi
Head of the Department
Department of Electronics and Computer Engineering,
Institute of Engineering,
Tribhuvan University,
Pulchowk, Nepal.

ACKNOWLEDGMENT

The credit of this thesis goes out to many people. The project would not have been conducted smoothly without their support.

Firstly I would like to thank the MSc department of providing me opportunity to conduct this thesis work. I am thankful to MSc. Program Coordinator Mr. Sharad K. Ghimire and HOD Mr. Shashidhar Ram Joshi for supporting me and this project. I am indebted to my supervisor Assist. Prof. Daya Sagar Baral for his continuous support and valuable suggestions.

I also want to thank all the persons that are directly or indirectly involved in the process of thesis writing.

Manish Shrestha
(063/MSI/609)

ABSTRACT

Biometric systems have been researched intensively for security issues. Biometric systems can uniquely identify a particular identity. Among the biometric systems face recognition system is one of the most popular. In this approach the individuals are identified by the feature of face. Research has been in progress since 1980's with numerous applications henceforth. Currently, many face recognition applications are available commercially for criminal identification, security system, image processing etc.

Face recognition is a popular research area where there are different approaches studied in the literature. The goal of face recognition system is straightforward; Compare the captured images with images stored in database and recognize the faces already stored in database. In this thesis, a holistic Principal Component Analysis (PCA) based method, namely Eigenface method, Linear Discriminator Analysis (LDA) based method, namely Fisherfaces and Independent Component Analysis (ICA) based method are studied in detail. These algorithms are studied in detail and these three methods are compared.

TABLE OF CONTENTS

Chapter 4
Simulations

Chapter 5
Results

Chapter 6
Conclusions

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background

Face recognition has been an active research area over the last 30 years. There is a great interest of many researchers on the face recognition problem. Among these researchers are the engineers, neuroscientists, and psychophysicists studying this popular problem in different fields and in different points of view. Psychologists and neuroscientists mainly deal with the human perception part of the topic, whereas engineers studying on machine recognition of human faces deal with the computational aspects of face recognition.

The topic of face recognition for real-world environments has garnered tremendous attention. There are several application areas of face recognition in our real life such as identification of personnel using credit cards, passport checks, entrance control, criminal investigations, etc For example, the ability to model a particular face and distinguish it would vastly improve security systems. Also, detecting faces allows further enhancement and noise reduction on the image.

The necessity for personal identification in the fields of private and secure systems made face recognition one of the main fields among other biometric technologies. Biometric systems have been researched intensively for security issues. Biometric systems can uniquely identify a particular identity. Among the biometric systems face recognition system is one of the most popular. The importance of face recognition rises from the fact that a face recognition system does not require the cooperation of the individual while the other systems need such cooperation.

Face recognition has gained much attention in recent years and has become one of the most successful applications of image analysis and understanding. A general statement of the problem can be formulated as follows: Given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces. Currently, image based face recognition techniques can be divided into two groups based on the face representation which they use:

1. **Appearance Based** which use holistic texture features and are applied to either whole-face or specific regions in a face image, and

2. **Feature Based** which use geometric facial features (mouth, eyes, brows, cheeks etc.) and geometric relationships between them.

Among many approaches to the problem of face recognition, the appearance-based subspace analysis provides the most promising results. Subspace analysis is done by projecting an image into a lower dimensional space (subspace) and after that recognition is performed by measuring the distances between known images and the image to be recognized. The most challenging part of such a system is finding an adequate projection.

The topic seems to be easy for a human, where limited memory can be a main problem; whereas the problems in machine recognition are manifold. Even with pre-processing the dynamic nature of face images presents significant problems during the recognition process. A face recognition system may be graded "robust" or "weak" based on its recognition performance under different situations of the face images. A robust system is capable of recognizing an image under different variations. Some of possible problems for a machine face recognition system are mainly:

1. Facial Expression: Face images of the same person can differ in expressions when smiling or laughing.

2. Illumination: Face images of the same person can be taken under different illumination conditions such as, the position and the strength of the light source can be modified.

3. Aging: Face images also vary greatly with age. A picture just a few months apart may not look alike at all.

4. Size of image: The same face can be presented to the system at different scales. This may happen due to the focal distance between the face and the camera. As this distance gets closer, the face image gets bigger.

5. Frontal vs Profile: The same face can be presented to the system at different perspectives and orientations. For instance, face images of the same person could be taken from frontal and profile views. Besides, head orientation may change due to translations and rotations.

6. Noise: A robust face recognition system should be insensitive to noise generated by frame grabbers or cameras. Also, it should function under partially occluded images.

1.2 Role of Biometrics

The information and intellectual property are under seize from many unauthorized personnel. This has resulted in adoption of more secure authentication methods for user access in various areas such as banks, post office, airports, automatic teller machines etc. Conventional password based systems such as, token, ID cards can be easily breached by others. Such requirement for reliable personal identification in access control has resulted in an increased interest in biometrics.

The International Biometrics Group (IBG) [33] defines biometrics simply as "the automated use of physiological or behavioural characteristics to determine or verify identity." Biometrics was a monotonous science relying only on fingerprint identification. However, over decades biometrics has become a multidisciplinary science representing millions of dollars. The various biometric systems used for identification are: Signature verification, finger print Recognition, Iris Recognition, Voice Recognition, Face Recognition etc.

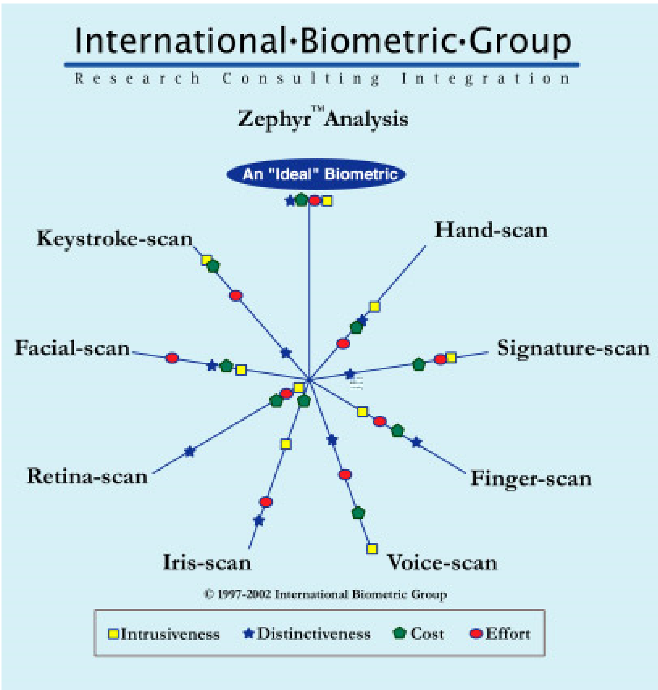

Fig 1.1: Comparison of various Biometric Features

These different biometric disciplines are appropriate for high security applications, however, cannot be used in general conditions. These techniques require people to position their body according to the system and pause for some time. Such cases demand systems which work well without the cooperation of people at any time. The increased security awareness has caused a tremendous amount of investment on the Biometric systems. Global2002 industry revenues of $601million are expected to reach $4 billion by 2007 [33].

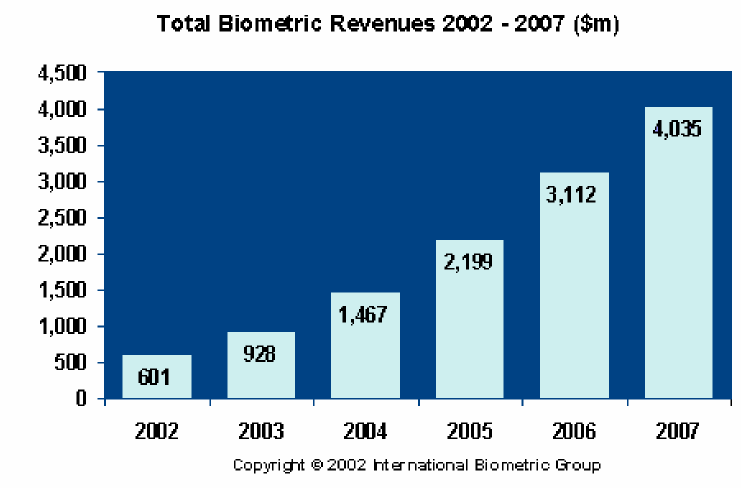

Fig 1.2: Global Biometric Revenues

As one of the most successful applications of image analysis and understanding, face recognition has recently received significant attention, especially during the past few years. A face recognition system would allow people to be identified by simply walking past a surveillance camera. The reasons for this trend are the wide range of commercial and law enforcement application and the availability of feasible technologies after years of research.

1.3 Scope of Thesis

In this thesis, the performances of three statistical face recognition techniques were studied on different face databases. The face recognition techniques Principal Component Analysis (PCA) namely Eigenfaces, Linear Discriminator Analysis (LDA) namely Fisherface and Independent Component Analysis (ICA) were studied in detail. In all three algorithms the PCA algorithm is a must. It is applied in other two algorithms LDA and ICA as well.

All three methods are implemented on the Olivetti Research Laboratory (ORL) face database and the Yale Face Database. There was also use of other test images as well. Experiments are conducted for relative performance evaluations. The three projection methods and their accompanied distance metrics are compared in completely equal working conditions.

1.4 Thesis Outline

The rest of the paper is organized as follows. Chapter 2 gives a brief description of the pattern recognition problem. This is closely related to face recognition and face recognition evolved from pattern recognition.

The Chapter 3 gives description of the algorithms to be compared. This section elaborates the basis of face recognition. The thesis is focused to provide a proper solution to the problem of face recognition. This section elaborates the different techniques which have been used and which are under research.

In Chapter 4 the details of experimental design are explained.. After considering the proper algorithm data are required for tests. The data in this case are images, which must be properly selected for maximum efficiency. The data and their requirements regarding the algorithms are described.

Chapter 5 reports the results and compares it to other research groups. The results are analyzed. The performance analysis of the algorithms PCA, ICA and LDA coupled with proper metric is shown. The results are verified with the claims of different researchers. The results indicate whether there is a proper algorithm for face recognition or not.

Chapter 6 concludes the paper.

CHAPTER 2: PATTERN RECOGNITION AND FACE RECOGNITION

2.1 Pattern Recognition Overview

The object detection and recognition problem is one of the most important research areas in pattern recognition and computer vision. One of the major problems in the design of modern information systems is automatic pattern recognition.

Humans can identify patterns or regularities in a set of observations. From the early development of computers, scientists and engineers tried to imitate this ability by mechanical means, either partially or in its entirety [5]. A human being is a very sophisticated information system, partly because he possesses inherent pattern recognition capability.

Recognition of concrete patterns by human beings may be considered as a psychophysiological problem [30] which involves a relationship between a person and a sensory stimulus. Human recognition is in reality an estimation of the relative odds that a data can be associated with one of a set of known data which depend on our past experience and which form the clues which gives insight information for recognition. Thus, the problem of pattern recognition may be regarded as one of discriminating the input data between populations via the search for features or invariant attributes among members of a population.

Pattern recognition is a process of categorizing any sample of measured or observed data as a member of one of the several classes or categories. A Pattern class can be defined as a set of patterns that share some properties in common. Since patterns in the same class share some properties in common, they can be easily differentiated. Pattern can be defined as a quantitative or structural description of an object or some other entity of interest.

A pattern is the description of an object. According to the nature of the patterns to be recognized, recognition acts can be divided into two major types [30]:

Recognition of concrete items. This may be referred to as sensory recognition, which includes visual and aural pattern recognition. This recognition process involves the identification and classification of spatial and temporal patterns. Examples of spatial patterns are characters, fingerprints, physical objects, and images. Temporal patterns include speech waveforms, time series, electrocardiograms and target signatures.

Recognition of abstract items. On the other hand, an old argument, or a solution to a problem can be recognized. This process involves the recognition of abstract items and can be termed conceptual recognition.

2.2 Outline of a Typical Pattern Recognition System

In Figure 2.1, functional block diagram of an adaptive pattern recognition system is shown. The functional breakdown provides a clear picture for the understanding of the pattern recognition problem.

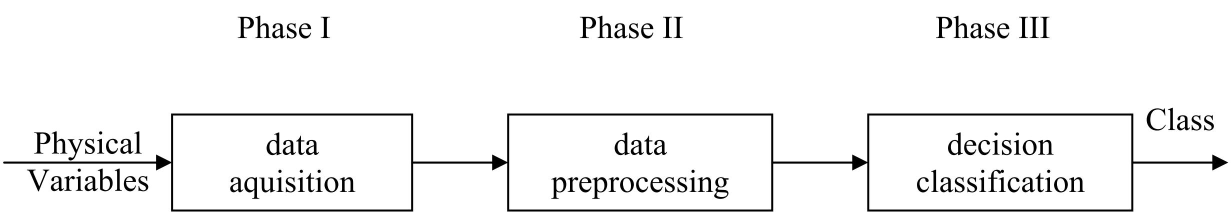

Fig 2.1: Conceptual representation of PR system

Correct recognition will depend on the amount of discriminating information contained in the measurements and the effective utilization of this information. In some applications, background information is essential in achieving accurate recognition. A flexible pattern recognition system is resistant to deviations and distortions in the information.

A pattern recognition task can be divided into three steps: data acquisition, data preprocessing and data classification. In the data acquisition phase analog data form the physical world are gathered. Here, the physical variables are converted into measured data. The data preprocessing step consists of a process of feature extraction. The reason for including feature extraction is because the amount of data obtained through the data acquisition phase is tremendous and must be reduced to a manageable amount while still carrying discriminatory features. The third phase actually is a classifier that is in a form of set of decision functions.

2.3 Training and Learning

The development of a PR application starts with the evaluation of the type of features to be used and the adequate PR approach for the problem at hand. For this purpose an initial set of patterns is usually available [31]. When a complete set of knowledge about the patterns to be recognized is available, the decision function may be determined with precision of the basis of this information. The performance of a PR system is usually evaluated in terms of error rates for all classes and an overall error rate. With this evaluation on a set of data we obtain patterns.

The learning and training takes place only during the design of the pattern recognition system. Once acceptable results have been obtained with the training set of patterns, the system is applied on the performing recognition on samples from the real time environment. The quality of the recognition is determined by how closely the training patterns resemble the actual data with which the system will be confronted during normal operations

In order to obtain better estimates of a Pattern Recognition system performance it is indispensable to evaluate it using an independent set of patterns. The independent set of patterns is called a test set. Test set estimates of a PR system performance give us an idea of how well the system is capable of generalizing its recognition abilities to new patterns.

2.4 Supervised and Unsupervised Pattern Recognition

To classify a pattern into a category is itself a learning process. The pattern classification system should have the ability to learn and to improve its performance of the classification through learning. The improvement in the system takes time with proper learning process [32].

In most cases, representative samples from each class under consideration are available. In these situations, supervised pattern recognition techniques are applicable. In a supervised learning environment, the system is taught to recognize patterns by means of various adaptive learning schemes. The essentials of this approach are a set of training samples of known classification and the implementation of an appropriate learning procedure. In some applications, only a set of training patterns of unknown classification may be available. In these situations, unsupervised pattern recognition techniques are applicable. As mentioned above, supervised pattern recognition is characterized by the fact that the correct classification of every training pattern is known. The advantage of using supervised learning system to perform the pattern classification is that it can construct a linear or a nonlinear decision boundary between different, and therefore offer a practical method for solving highly complex pattern classification problems.

Many cases have no a priori knowledge of categories into which the patterns are to be classified .In the unsupervised case; one is faced with the problem of actually learning the pattern classes present in the given data. This problem is also known as "learning without a teacher". In unsupervised learning, patterns are associated by themselves into clusters based on some properties in common. These properties are sometimes known as features.

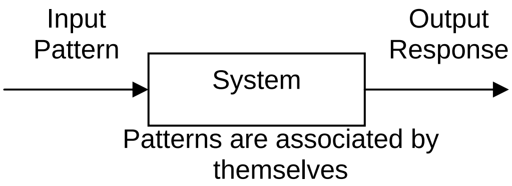

Fig 2.2: Unsupervised Learning

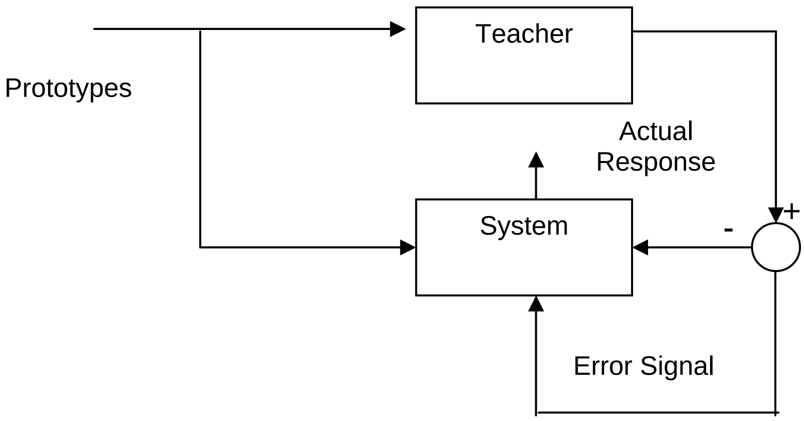

Fig 2.3: Supervised Learning

2.5 Face Recognition

Face recognition is a pattern recognition task performed specifically on faces. It can be described as classifying a face either "known" or "unknown", after comparing it with stored known individuals. It is also desirable to have a system that has the ability of learning to recognize unknown faces.

The detailed aspects of Face Recognition can be found in Chapter 3.

CHAPTER 3: SURVEY ON FACE RECOGNITION METHODS

3.1 Face Recognition Methods

The face is our primary focus of attention in social intercourse, playing a major role in conveying identity and emotion. The ability to deduce intelligence or character from facial appearance is the theme where, the human ability to recognize faces is remarkable. Humans can recognize thousands of faces our lifetime and identify familiar faces at a glance even after years of separation. This skill is quite robust, despite large changes due to viewing conditions, expression, aging, and distractions such as glasses, beards or changes in hair style.

Human often use faces to recognize individuals and advancements in computing technology over the past few decades have enabled similar recognitions automatically. A human can easily detect and identify an individual whether it is viewed from front side or even images which are distorted. However, the same is not true for a computer. The area to be recognized must be detected then only the recognition is possible. The images required for recognition by computer must be taken in a controlled environment so that the recognition becomes easy. Also, the real world pictures are not according to the specifications required by the recognition algorithm with various tilting, lighting conditions, camera resolution issues etc. These images are manually reproduced by different preprocessing techniques like cropping, rotation, histogram equalization and masking.

3.1.1 Human Face Recognition

The basis of all artificial face recognition systems is the human face recognition. Artificial face recognition simply tries to mimic the different concepts that the human brain may use. Sitting on top of your shoulders is the most sophisticated computer ever developed in the history of the known universe. It is not at all obvious how faces are encoded or decoded by the human brain. Attempts to replicate its working have been in progress for about three decades. Unfortunately developing a computational model of face recognition is quite difficult, because faces are complex, multi-dimensional visual stimuli. The human face recognition system not only uses the 2D or 3D visual information [8] on the face. There are information regarding different senses: visual, auditory, tactile, etc. All the information from these senses is used either individually or collectively for the identification of human faces.

There are also various other issues regarding the face recognition. It has been found that the face is not the only feature studied in human face recognition. There are also issues such as the background, hair, expressions, emotions, attractiveness and unattractiveness of the face [8]. Humans perceive other faces of other humans based on the above mentioned characteristics as well. It has also been found that the upper part of the face is more useful than the lower part of the face for recognition. Also, visual attributes (e.g. beauty, attractiveness, pleasantness, etc.) play an important role in face recognition. Moreover, human face recognition is much sharper while regarding these characteristics rather than the face itself.

Both holistic and feature information are important for the human face recognition system. Studies suggest the possibility of global descriptions serving as a front end for better feature-based perception [8]. If there are dominant features present such as big ears, a small nose, etc. holistic descriptions may not be used. A caricature can refer to a portrait that exaggerates or distorts the essence of a person or thing to create an easily identifiable visual likeness. In literature, a caricature is a description of a person using exaggeration of some characteristics and oversimplification of others. A caricature does not contain as much information as a photograph; however, it manages to capture the most important characteristics of the face images

For humans, photographic negatives of faces are difficult to recognize. But, there is not much study on why it is difficult to recognize negative images of human faces. Also, a study on the direction of illumination [17] showed the importance of top lighting; it is easier for humans to recognize faces illuminated from top to bottom than the faces illuminated from bottom to top.

3.1.2 Machine Face Recognition

The studies on Machine Face Recognition started around three decades ago. Since the early 1990's, research interest on machine recognition of faces has grown tremendously. The reasons may be:

- Availability of Real time hardware
- The growing need of surveillance applications
- The studies on real time computation
- An increase in civilian and commercial projects

For a machine face recognition the input to the system is an unknown face, and the system reports back the determined identity from a database of known individuals. The basic question for this purpose is what method of encoding of the face should be taken to achieve face recognition. The two major approaches are: Appearance based approach and Feature based approach. Besides these methods currently to improve the chances of recognition hybrid methods based on the combination of different other face recognition approaches are also used to overcome the shortcomings of different approaches.

All face recognition algorithms have to identify a distinct pattern for recognition. The algorithms extract a structural characteristic based on the picture which is unique to each image. Research in the field primarily intends to generate sufficiently reasonable patterns of human faces so that the face can be identified.

The question with all recognition methods is: what those patterns are? A general structure of the recognition system is shown below.

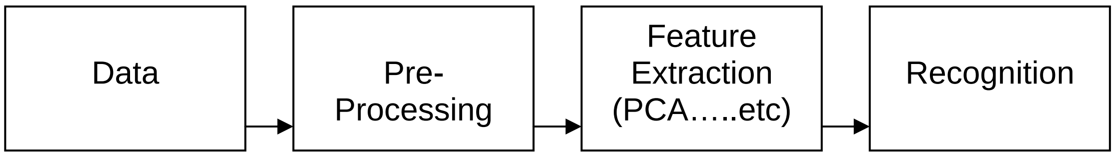

Fig 3.1: Basic Stages of Recognition Process

3.1.2.1 Feature Based Approach

The feature based approach is one of the oldest methods which still give satisfactory results. Faces are differentiated so that the different local features can be retrieved for identifying faces. The information is generally derived from different features on the face such as eye, nose, mouth, hair, etc. The distance information is then subjected to standard statistical pattern recognition techniques and/or neural network approaches are employed for recognition purposes. One of the well known geometrical-local feature based methods is the Elastic Bunch Graph Matching (EBGM) technique.

3.1.2.2 Appearance Based Approach

The appearance based approach is also called the subspace analysis method. This method first reduces the given image a high dimensional data into a low dimensional data; thus, it is called subspace analysis method. The low dimensional representation [18] is one of the key to the success of the feature based approach. The primary reasons why low dimensional image are:

- Handling the high dimensional data is computationally very expensive.

- For appearance based approach the number of parameters required for identification grows with dimensionality

- Also for feature based approach, the number of images required for training properly is also properly is fairly high.

The main effort for this approach is to find a proper subspace such that reduces the dimension of the image properly while preserving enough data for the analysis afterward. The latter processes are nearly the same, which includes pattern recognition techniques to analyze the extracted features of the image. Techniques such as PCA, LDA and ICA are used to extract the features of the image.

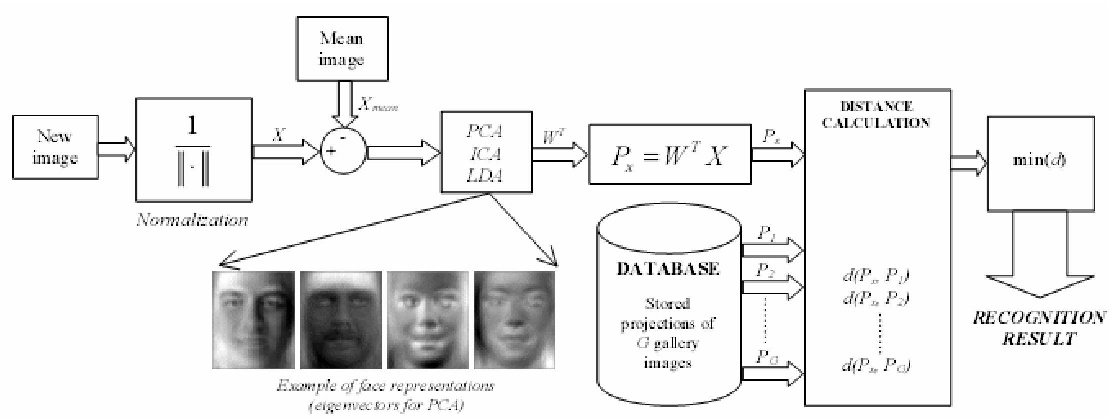

Fig 3.2: A general subspace face recognition system

3.2 Principal Component Analysis

PCA is a statistical dimensionality reduction method, which produces the optimal linear least squares decomposition of a training set. Kirby and Sirovich [1] applied PCA to representing face and Turk and Pentland [2] extended PCA to recognizing faces. It is also known as eigenspace projection or Karhunen Loeve Transformation.

In mathematical terms, the principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images, treating an image as point (or vector) in a very high dimensional space is sought. The eigenvectors are ordered, each one accounting for a different amount of the variation among the face images. PCA evaluates subspace whose basis vectors correspond to the maximum variance direction in the original image space.

Sample face images and the corresponding eigenfaces are shown in Figure 3.3 and in Figure 3.5 respectively. Each eigenface deviates from uniform gray where some facial feature differs among the set of training faces. Eigenfaces can be viewed as a sort of map of the variations between faces.

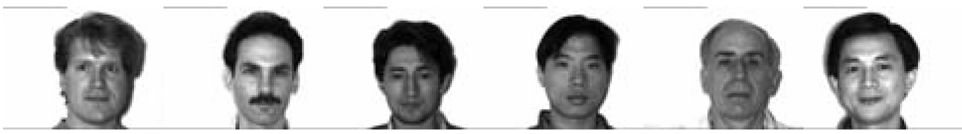

Fig 3.3: A subset of training set face images.

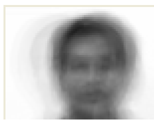

Fig 3.4: Average Face of the training set

Fig 3.5: Some eigenfaces

Each individual face can be represented exactly in terms of a linear combination of the eigenfaces. Each face can also be approximated using only the "best" eigenfaces, those that have the largest eigenvalues, and which therefore account for the most variance within the set of face images.

The thesis follows the method which was proposed by M. Turk and A. Pentland [2] in order to develop a face recognition system based on the eigenfaces approach. They argued that, if a multitude of face images can be reconstructed by weighted sum of a small collection of characteristic features or eigenpictures, perhaps an efficient way to learn and recognize faces would be to build up the characteristic features by experience over time and recognize particular faces by comparing the feature weights needed to approximately reconstruct them with the weights associated with known individuals.

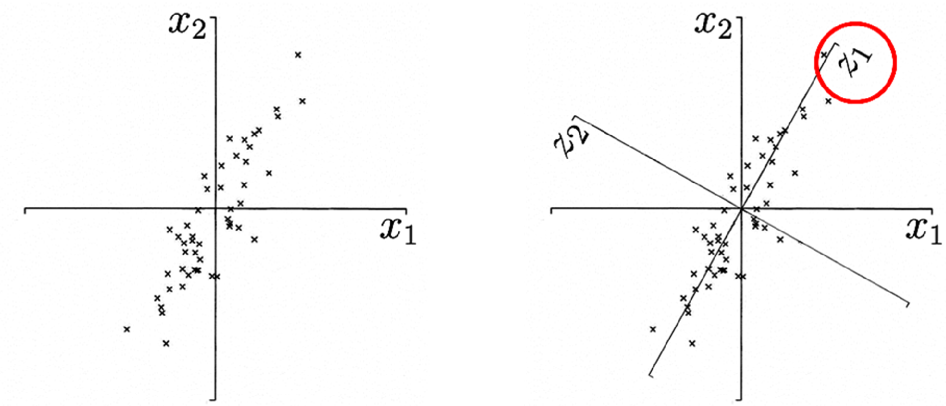

Fig 3.6: Concept of PCA

(Left Fig shows the data and the right fig shows the new basis for representation)

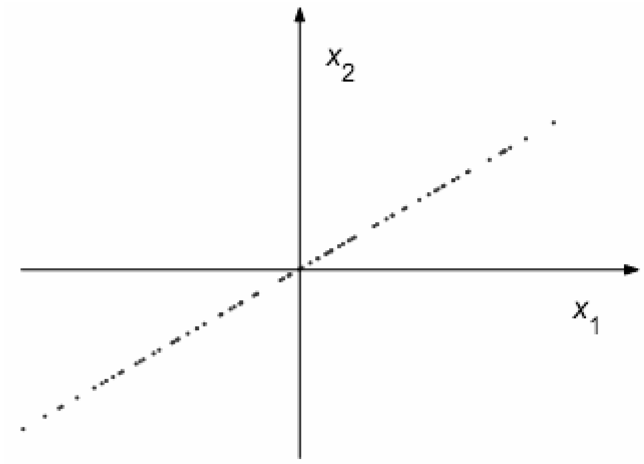

Fig 3.7: PCA reduction in 1D

(Data reduced in new basis axes)

Eigenvectors can be considered as the vectors pointing in the direction of the maximum variance and the value of the variance the eigenvector represents is directly proportional to the value of the eigenvalue Hence, the eigenvectors are sorted with respect to their corresponding eigenvalues. The eigenvector having the largest eigenvalue is marked as the first eigenvector, and so on. In this manner, the most generalizing eigenvector comes first in the eigenvector matrix. PCA evaluates subspace whose basis vectors correspond to the maximum variance direction in the original image space.

3.2.2 Eignevalues and Eigenvectors

An x * y matrix A is said to have an eigenvector X, and corresponding eigenvalue λ if

$$A*X = \lambda X \tag{3.1}$$

Evidently, Eq. (3.1) can hold only if

$$\text{Det } |A- \lambda I| = 0 \tag{3.2}$$

which, if expanded out, is an Nth degree polynomial in λ whose root are the eigenvalues. This proves that there are always N (not necessarily distinct) eigenvalues. Equal eigenvalues coming from multiple roots are called "degenerate".

A matrix is called symmetric if it is equal to its transpose,

$$A = A^T \text{ or } a_{ij} = a_{ji} \tag{3.3}$$

It is termed orthogonal if its transpose equals its inverse,

$$A*A^T = A^T*A = I \tag{3.4}$$

finally, a real matrix is called normal if it commutes with is transpose,

$$A^TA = AA^T \tag{3.5}$$

Theorem: Eigenvalues of a real symmetric matrix are all real. Contrariwise, the eigenvalues of a real nonsymmetric matrix may include real values, but may also include pairs of complex conjugate values. The eigenvalues of a normal matrix with nondegenerate eigenvalues are complete and orthogonal, spanning the N-dimensional vector space.

The success of Eigenfaces is based on the evaluation of the eigenvalues and eigenvectors of the real symmetric matrix A^TA that is composed from the training set of images. Root searching in the characteristic equation, Eqn. (3.2) is usually a very poor computational method for finding eigenvalues. During the programming phase of the above algorithm, a more efficient method [2, 3, 4] was used in order to evaluate the eigenvalues and eigenvectors. At first, the real symmetric matrix is reduced to tridiagonal form with the help of the "Householder" algorithm. The Householder algorithm reduces an N*N symmetric matrix A to tridiagonal form by N-2 orthogonal transformations. Each transformation annihilates the required part of a whole column and whole corresponding row. After that, eigenvalues and eigenvectors are obtained with the help of QR transformations. The basic idea behind the QR algorithm is that any real symmetric matrix can be decomposed in the form A = QR where Q is orthogonal R is upper triangular. The workload in the QR algorithm is $O(N^3)$ per iteration for a general matrix, which is prohibitive. However, the workload is only O(N) per iteration for a tridiagonal matrix, makes it extremely efficient.

3.2.2 Calculating Eigenfaces

We will assume that M sample images are being used. Each sample image will be referred to as $\tilde{\mathbf{A}}_n$ where the subscript indicates the corresponding n^{th} sample image ($1 \le n \le M$). Each $\tilde{\mathbf{A}}_n$ should be a column vector. Generally images are thought of as a matrix of pixels. Converting this to a column form is a matter of convenience, it can be done in either column or row major form, so long as it is done consistently for all sample images it will not affect the outcome. The size of the resulting $\tilde{\mathbf{A}}_n$ column vector will depend on the size of the sample images. If the sample images are x pixels across and y pixels tall, the column vector will be of size (x*y), so that a typical image of size 256 x 256 becomes a vector of dimension 65,536. These original image sizes must be remembered if one wishes to view the resulting eigenfaces, or projections of test images into face-space. This allows a normal image to be constructed from a column vector of image pixels.

The training set of face images are $\tilde{\mathbf{A}}_1$, $\tilde{\mathbf{A}}_2$,..., $\tilde{\mathbf{A}}_M$. The average image **Ø**, is to calculated, as follows:

$$Ø = \frac{1}{M}\sum_{i=1}^{M}\tilde{A}_i \qquad (3.6)$$

This average image will be a column vector of the same size as the sample images x*y. When the vector is interpreted as a normal image it is achieved as shown in the figure 4 above.

The next step is to calculate the difference faces by subtracting the average face from each sample image. Each face differs from the average by the vector as below:

$$\ddot{O}_n = \tilde{A}_n - \emptyset, \text{ where } 1 \leq n \leq M \tag{3.7}$$

Each will be a column vector the same size as our sample image vectors x*y. The purpose of calculating these difference faces is to allow us to calculate the covariance matrix for our sample images. The covariance matrix is defined by AA^T where A = [$\ddot{O}_1$ $\ddot{O}_2$ $\ddot{O}_3$……$\ddot{O}_n$], that is the columns of the A matrix are formed by the differences faces $\ddot{O}_n$. The matrix A will be of size (x*y) × M.

The M orthonormal vectors , $X_{n,}$ which best describes the distribution of the data are required. The kth vector, X_k, is chosen such that

$$\lambda_k = \frac{1}{M} \sum_{n=1}^{M} (X_k^T \emptyset_n)^2 \tag{3.8}$$

is maximum subject to

$$X_l^T X_k = \delta_{lk} = \begin{cases} 1, & \text{if } l = k \\ 0, & \text{otherwise} \end{cases} \tag{3.9}$$

The vectors X_k and scalars λ_k are the eigenvectors and eigenvalues, respectively of the covariance matrix A

$$C = \frac{1}{M} \sum_{n=1}^{M} \ddot{O}_n \ddot{O}_n^T = AA^T \tag{3.10}$$

The covariance matrix is, however, of size (x*y)*(x*y) and determining the (x*y) eigenvectors and eigenvalues is an intractable task for typical image sizes. A computationally feasible method to find these eigenvectors is required. For example, a standard sample image might be approximately 200x200 pixels. Clearly, doing these calculations on the resulting matrix of size 40000 × 40000 is going to be strenuous. A computationally practical method for evaluating the eigenvectors for such prescribed data is required.

If the number of image space is less than the dimension of the image ($M < (x*y)^2$), there will be only M-1, rather than $(x*y)^2$, meaningful eigenvectors. The remaining eigenvectors will have associated eigenvalues of zero. The $(x*y)^2$ dimensional eigenvectors in this case can be solved by first solving the eigenvectors of an M x M matrix such as solving 16 x 16 matrix rather than a 16,384 x 16,384 matrix and then, taking appropriate linear combinations of the face images $Ö_i$. This is verified as below:

Consider the eigenvectors v_i and eigenvalues μ_i of A^TA such that

$$A^TAv_i = \mu_i v_i \tag{3.11}$$

Pre-multiplying both sides by A,

$$AA^TAv_i = \mu_i Av_i \tag{3.12}$$

Here, Av_i are eigenvectors of $C= AA^T$. Now M*M matrix A^TA is constructed and the M eigenvectors v_i is found.

With this analysis, the calculations are greatly reduced, from the order of the number of pixels in the images $(x*y)^2$ to the order of the number of images in the training set M. In practice, the training set of face images will be relatively small $M<<(x*y)^2$, and the calculations become quite manageable. The associated eigenvalues allow us to rank the eigenvectors according to their usefulness in characterizing the variation among the images. The eigenvectors determine linear combinations of the M training set face images to form eigenfaces u_i.

$$u_k = \sum_{l=1}^{M} X_{lk} \text{Ö}_l \tag{3.13}$$

Some of the evaluated eigenfaces are illustrated in figure 3.5

The face subspace is quite interesting as it is able to use a few values and represent an entire image. A 400x400 pixel picture may be represented by few values which may be as less as 3. The reduction in dimension removes the information that are not useful and correlates the images into orthogonal components.

3.2.3 Recognition of Faces

Now that the eigenfaces have been created, they must be uses in order to recognize or analyze images. This is a very simple procedure, as illustrated by the following formula.

$$\mathbf{\grave{U}} = U^T (\mathbf{\tilde{A}_n} - \mathbf{Ø}) \tag{3.14}$$

Put simply, the vector of weights is found by multiplying the transpose of the matrix U^T by a vector that is found by subtracting the average face image (Ø, a column vector) from a sample or test image ($\tilde{A}_n$, a column vector). It should be noted that although $\tilde{A}_n$ represents the nth sample image in our nomenclature, this image could be any sample or test image, as long as it has already been converted into a column vector.

The above operation may also be carried out one weight one a time. The following formula realizes this

$$\grave{u} = u_n^T (\mathbf{\tilde{A}_n} - \mathbf{Ø}) \tag{3.15}$$

Here, each $ù_n$ is calculated by considering the n^{th} eigenface (a column vector) as well as the average image and image to be projected into face space. This process is repeated for n = 1, 2, …k so that the weights are calculated for each eigenface.

Now eigenvectors (PCA basis vectors) define a subspace of the face images called face space. u_k represents the projections of the known faces on the face space. To identify an unknown image, the image projections ù is compared to the weights of known faces.

3.3 Linear Discriminator Analysis

LDA is also known as Fisher's Discriminant Analysis and it searches for those vectors in the underlying space that best discriminate among classes. The basic idea of LDA is to find a linear transformation such that feature clusters are most separable after the transformation which can be achieved through scatter matrix analysis

LDA is also closely related to principal component analysis (PCA) and factor analysis in that both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities.

LDA searches for the best projection to project the input data, on a lower dimensional space, in which the patterns are discriminated as much as possible. For this purpose, LDA tries to maximize the scatter between different classes and minimize the scatter between the input data in the same class.

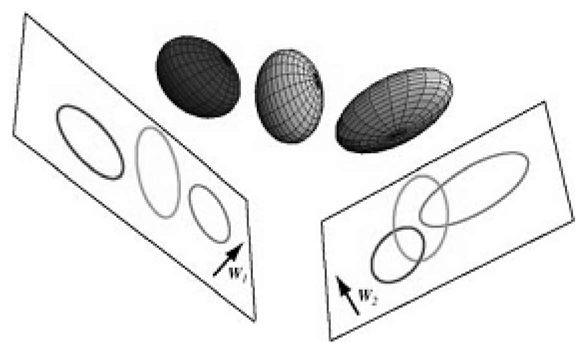

Fig 3.8: LDA Concept (Distributions projected into 2 dimensional subspaces represented by w1 and w2. Here w1 is optimal for LDA)

3.3.1 Fisherfaces calculations

As in the case of eigenfaces, the training set of face images is $\tilde{A}x$. Then the mean image is evaluated as:

$$Ø = \frac{1}{M}\sum_{i=1}^{M}\tilde{A}_i \tag{3.16}$$

In LDA, mean face images are also calculated for each face class; this is due to need for the calculation of each face classes inner variation. Hence, for each of c individuals having qi training images in the database.

$$Øc_i = \frac{1}{qi}\sum_{i=1}^{qi}\tilde{A}_i \tag{3.17}$$


(a) (b)


(c) (d)

Fig 3.9: 4 of the 15 different classes available in database

LDA creates a linear combination of these which yields the largest mean differences between the desired classes. Mathematically speaking, for all the samples of all classes, we define two measures: 1) one is called within-class scatter matrix, 2) the other is called between-class scatter matrix

$$S_W = \sum_{i=1}^{c}\sum_{x_k \in X_i}(\tilde{A}_k - Ø_i)(\tilde{A}_k - Ø_i)^T \tag{3.18}$$

$$S_B = \sum_{i=1}^{c} M_i.(Ø_i - Ø).(Ø_i - Ø)^T \tag{3.19}$$

where M_i is the number of training samples in class i, c is the number of distinct classes.

Linear Discriminant Analysis can not deal with the problem of one training image per class case; so M_i should be always greater than 1 In that case, S_w turns out to be an identity matrix; hence the solution reduces to a standard Eigenface approach with S_w being the covariance matrix [15].

LDA creates a linear combination of these which yields the largest mean differences between the desired classes. Mathematically speaking, for all the samples of all classes, two measures are defined: 1) one is called within-class scatter matrix, 2) the other is called between-class scatter matrix

Now, a distance measure is required that gives a maximum separation between the classes. However, the distance between projected means is not a very good measure since it does not take into account the standard deviation within the classes

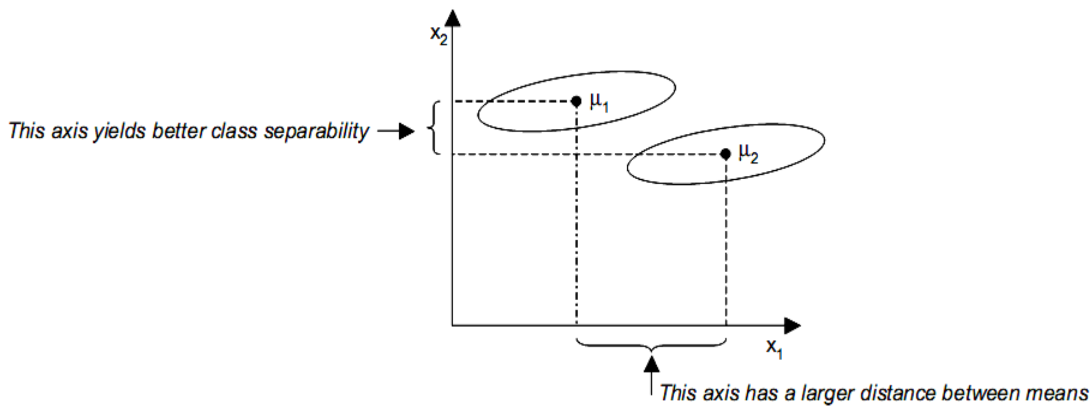

Fig 3.10: Relative separation between the classes on different projections

The solution proposed by Fisher is to maximize a function that represents the difference between the means, normalized by a measure of the within-class scatter. The Fisher linear discriminant is defined as the linear function $w^T x$ that maximizes the criterion function

$$J(w) = \frac{|\mu^1 - \mu^2|^2}{s1^2 + s2^2} \tag{3.20}$$

This function allows us to look at the projections such that they are projected very close to each other but as far as possible.

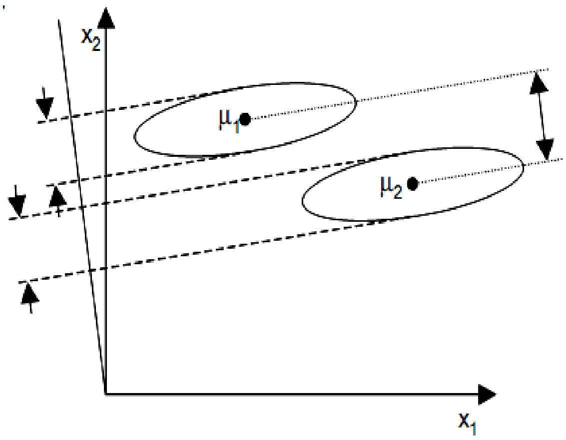

Fig 3.11: Fishers projection

The Fishers criterion can be expressed in terms of S_W and S_B as:

$$J(W) = \frac{w^T S_B w}{w^T S_W w} \tag{3.21}$$

With this new scatter whose projections give the highest distance between the classes; an optimal projection matrix W must be chosen. The columns of W are the eigenvectors corresponding to the largest eigenvalues of the following generalized eigenvalue problem.

$$W = [w_1 \mid w_2 \mid ...... \mid w_{c-1}] = \arg\max \left\{ \frac{w^T S_B w}{w^T S_W w} \right\} \Rightarrow (S_B - \lambda_i S_W) wi = 0 \qquad (3.22)$$

The generalized eigenvalue problem on the right side of equation 2.17 can be simplified as.

$$S_W^{-1} S_B w = \lambda w \qquad (3.23)$$

This ratio is maximized when the column vectors of the projection matrix (W_{LDA}) are the eigenvectors of $S_W^{-1}S_B$. The projections of the images are obtained using the eigenvalues and eigenvectors obtained from equation 2.18. The projections obtained and be used implicitly for the purpose of face recognition

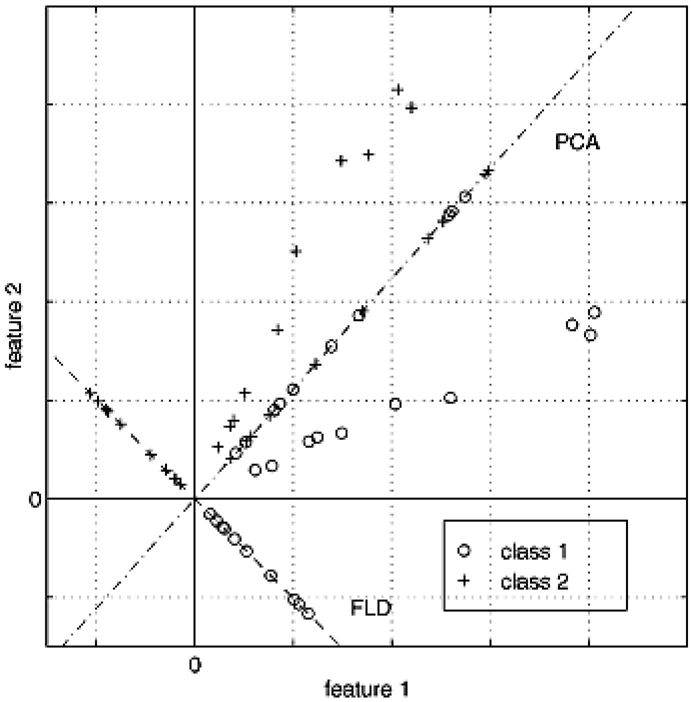

Fig 3.12: Comparison of Principal component analysis (PCA) and Fisher's Linear discriminant (FLD)

The figure above is a comparison of PCA and FLD or LDA for a two-class problem in which the samples from each class are randomly perturbed in a direction perpendicular to a linear subspace. For this example, N = 20; n = 2, and m = 1. So, the samples from each class lie near a line passing through the origin in the 2D feature space. Both PCA and FLD have been used to project the points from 2D down to 1D. Comparing the two projections in the figure, PCA actually smears the classes together so that they are no longer linearly separable in the projected space. It is clear that, although PCA achieves larger total scatter, FLD achieves greater between-class scatter, and, consequently, classification is simplified [22].

3.4 Independent Component Analysis

Independent Component Analysis (ICA) is a recently developed method in which the goal is to find a linear representation of nongaussian data so that the components are statistically independent; or as independent as possible. This captures the essential structure of the primary data. Independent Component Analysis (ICA) has emerged recently as one powerful solution to the problem of blind source separation [13].

Independent component analysis finds a set of directions in the data such that when the data points are projected onto these directions, the resulting data are statistically in-dependent (a much stronger condition that uncorrelated). Unlike PCA, these directions need not be orthogonal within the original space [1].

In a nutshell, the goal of ICA is the decomposition of a set of data in an a priori unknown linear mixture of a priori unknown source signals, relying on the assumption that the source signals are mutually statistically independent. This concept is in fact a fine-tuning of the well-known principal component analysis (PCA), where one aims at the decomposition in a linear mixture of uncorrelated components [23].

3.4.1 Blind Source Separation

Let us say there are three underlying source signals, and also three observed signals, denoted by $x_1(t)$, $x_2(t)$ and $x_3(t)$ the observed signals, which are the amplitudes of the recorded signals at time point t, and $s_1(t)$, $s_2(t)$ and $s_3(t)$ the original signals. The $x_i(t)$ are the weighted sums of the $s_i(t)$, where the weight coefficient indicate the change between the source and destination:

$$x_1(t)=a_{11}s_1(t)+a_{12}s_2(t)+a_{13}s_3(t) \tag{3.24}$$

$$x_2(t)=a_{21}s_1(t)+a_{22}s_2(t)+a_{23}s_3(t)$$

$$x_3(t)=a_{31}s_1(t)+a_{32}s_2(t)+a_{33}s_3(t)$$

The a_{ij} are constant coefficients that give the mixing weights. They are assumed unknown, since the values of a_{ij} cannot be known without knowing all the properties ok the system. The source signals are unknown as well; and the original signals from the mixtures $x_1(t)$, $x_2(t)$ and $x_3(t)$ must be calculated.. This is the blind source separation problem or cocktail party problem. Blind indicates that very little or anything is known about the original sources.

3.4.2 ICA Definition

ICA is a technique to separate linearly mixed [13] sources. ICA of a random vector consists of searching for a linear transformation that minimizes the statistical dependence between its components. The goal of ICA is to provide an independent image decomposition and representation. In other words, the goal is to minimize the statistical dependence between the basis vectors.

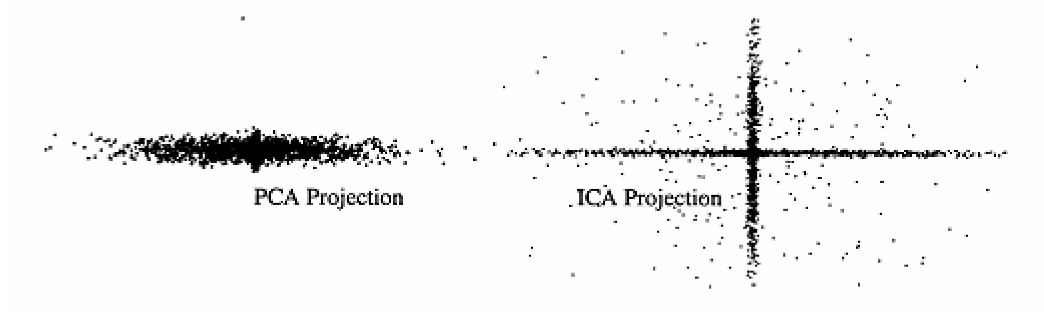

Fig 3.13: Comparison between PCA and ICA

Let s be the vector of unknown source signals and x be the vector of observed mixtures. If A is the unknown mixing matrix, then the mixing model is written as

$$x = As \tag{3.25}$$

It is assumed that the source signals are independent of each other and the mixing matrix A is invertible. Based on these assumptions and the observed mixtures, ICA algorithms try to find the mixing matrix A or the separating matrix W such that eqn 3.26 is an estimation [16] of the independent source signals.

$$u = Wx = WAs \tag{3.26}$$

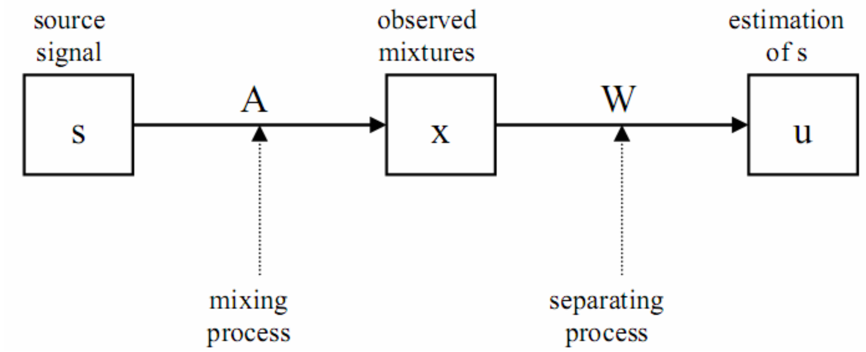

Fig 3.14: Blind source Separation Model

In case of PCA the training data was only decorrelated when the eigenvalues and eigenvectors of the difference matrix was evaluated. ICA performs whitening on the data so that there is docorrelation and unit variance on the data. Whitening is a useful preprocessing strategy in ICA, which means before application of the ICA algorithm, we transform the observed vector x linearly so that a new vector is obtained which is white, i.e. its components are uncorrelated and their variances equal unity.

The whitening transform [21] can be determined as $D^{1/2}R^T$, where D is the diagonal matrix of the eigenvalues and R is the matrix of orthogonal eigenvectors of the sample covariance matrix. This procedure evaluates the square root of the matrix data. ICA then goes one step further to transform the whitened data into a set of statistically independent signals.

A statistically independent signal is represented as:

$$f_u(u) = \Pi_i f_{u_i}(u_i) \tag{3.27}$$

where, f_u is the probability density function of u. Unfortunately, there may not be any matrix W that fully satisfies the independence condition, and there is no closed form expression [21] to evaluate W. For solution the W is approximated through iterative methods. Several different methods are available such as

InfoMax, Jade and FastICA. ICA algorithms recast the problem to iteratively optimize a smooth function whose global optima occurs when the output vectors *u* are independent. InfoMax, JADE and FastICA all maximize functions with the same global results. Thus, all three algorithms should converge to the same solution for any given data set. Zibulevsky and Pearlmutter test all three algorithms on a simulated blind-source separation problem, and report only small differences in the relative error rate: 7.1% for InfoMax, 8.6% for FastICA, and 8.8% for JADE [24].

Bartlett et al. provided two architectures [4] of ICA for face recognition task: Architecture I - statistically independent basis images, and Architecture II - factorial code representation. Here the Infomax algorithm proposed by Bell and Sejnowski [6] was chosen. InfoMax relies on the observation that independence is maximized when the entropy H(u) is maximized.

$$H(u) = -\int f_u(u) \log f_u(u) du \tag{3.27}$$

It gets its name from the observation that maximizing H(u) also maximizes the mutual information I(u,x) between the input and output vectors.

3.4.3 Architecture I statistically independent basis images

The goal in this approach is to find a set of statistically independent basis images. The input face images in X are considered to be a linear mixture of statistically independent basis images S combined by an unknown mixing matrix A. The ICA algorithm learns the weight matrix W, which is used to recover a set of independent basis images in the rows of U. Face image representations consist of the coordinates of these images with respect to the image basis defined by the rows of U.

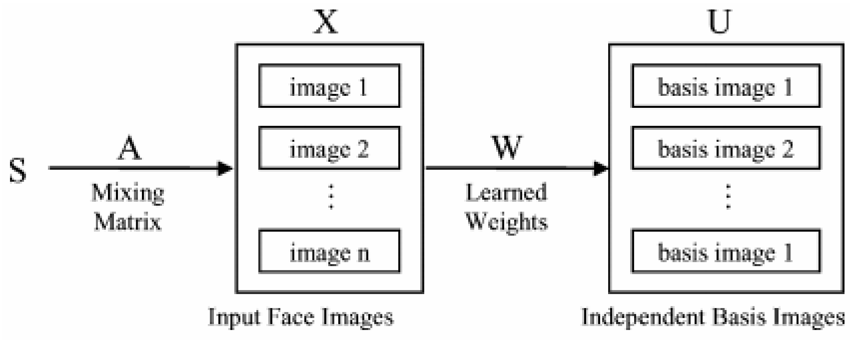

Fig 3.15: Statistically independent Basis Images

Let R be a *p*x*m* matrix containing the first m eigenvectors of a set of n face images. Let p be the number of pixels in a training image. The rows of the input matrix to ICA are variables and the columns are observations, therefore, ICA is performed on R^T. The m independent basis images in the rows of U are computed as $U = W * R^T$. Then, the *n*x*m* ICA coefficients matrix B for the linear combination of independent basis images in U is computed as follows:

Let C be the *n*x*m* matrix of PCA coefficients. Then

$$C = X * R \text{ and } X = C * R^T \quad (3.28)$$

From $U = W * R^T$ and the assumption that W is invertible t

$$R^T = W^{-1} * U \quad (3.29)$$

Therefore,

$$X = (C * W^{-1}) * U = B * U \quad (3.30)$$

Each row of B contains the coefficients for linearly combining the basis images to comprise the face image in the corresponding row of X. Also, X is the reconstruction of the original data with minimum squared error as in PCA.

3.4.4 Architecture II Factorial face code representation

While the basis images obtained in architecture I are statistically independent, the coefficients that represent input images in the subspace defined by the basis images are not. The goal of ICA in architecture II is to find statistically independent coefficients for input data. In this architecture, the input is transposed from architecture I, that is, the pixels are variables and the images are observation.

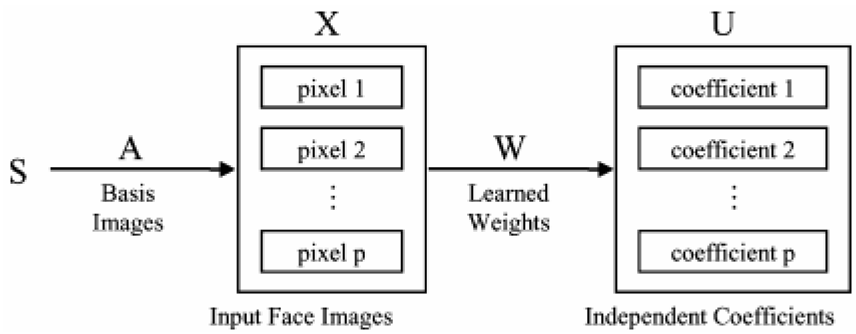

Fig 3.16: Statistically independent Coefficients

In this work, ICA is performed on the PCA coefficients rather than directly on the input images to reduce the dimensionality as in [4].

Chapter 4: Simulation

4.1 Image Data

The standard Yale data set [36] face images are used in the experiments. The original database was too large to be included in the experiment. A subset database was used which consists of 165 images in total. There were 15 subjects with 11 different images for each individual. The original image size was 320x243 pixels and grayscale. These images were resized to ¼th for simplicity in calculations. The ORL data set [37] was also considered for the purpose of testing the performance of Yale Data set. The ORL data set provided a set of 40 subjects with 10 different images for each individual.

4.2 Training

To train the Algorithms a subset of classes for which there are 11 images per class was used. To train the system 3 images per person is generally used. The number of training images such as 3 were taken under consideration as the operations could take too much time if large number of training images are used. The system was therefore trained with either 30 images when 10 different classes were considered. Thus, in general M=30 while c=10 were used. The remaining images were used in recognition stages. Thus the algorithms were trained roughly on 33% of the subjects later used the recognition stages. This allowed the exploration of the algorithms when subjected to different training conditions.

One thing to remember is that the PCA step was the key step in this process. All the three algorithms had to go through the PCA step. Firstly PCA is carried out on the database and the first representation was the eigenfaces. This subspace was used for recognition in PCA was used as the input for both LDA and ICA.

After all the subspaces have been derived, all the images form data sets were projected onto each subspace and recognition was performed.

4.3 Recognition

After the application of the PCA algorithm the problem of face recognition was now to a problem of pattern recognition where the patterns are indicated by the projections in which the training and the testing images are projected using the eigenfaces or the fisherfaces.

4.3.1 Euclidean Distance

It is the most common distance metric. The comparison of the projected image with training images can be evaluated with the Euclidean distance between the projections.

$$e_k = \sqrt{\sum_{i=1}^{n} (p_i - pk_i)^2} \tag{3.1}$$

Here p indicates the projection of the test image and p_k indicates the projection of the training images. The value e_k indicates the threshold. The value e_k basically evaluates the amount of error the new image has regarding the trained image

4.3.2 City Block Distance

Another distance measure for comparison between the projections is the city block distance. The city block distance between two points is defined as the sum of the absolute differences of their coordinates. It is also known as rectilinear distance or Manhattan distance as well [35].

$$e_k = \sum_{i=1}^{n} | p_i - pk_i | \tag{3.2}$$

4.3.3 Mahalanobis Distance

Mahalanobis distance is based on correlations between variables which different patterns can be identified and analyzed [35]. It is also called quadratic distance.

$$ek = \sqrt{(p_i - pk_i)^T S^{-1} (p_i - pk_i)} \tag{3.3}$$

where, S is the covariance matrix of projections

4.3.4 Cosine Angle Distance

The other common distance measures for recognition are Cosine angle. It measures similarity rather than distance or dissimilarity. Thus, higher value of Angular separation indicates the two objects are similar. The value of angular separation is [-1, 1] similar to cosine. It is often called as *Coefficient of Correlation.*

$$e_k = \frac{\sum_{i=1}^{n} (p_i * pk_i)}{(\sum_{i=1}^{n} p_i^{\,2} \sum_{i=1}^{n} pk_i^{\,2})^{\frac{1}{2}}} \tag{3.4}$$

CHAPTER 5: RESULTS

5.1 Tests

Different sets of experiments were conducted for the PCA, ICA and LDA algorithm. In those experiments, general performance of the algorithm, the effect of illumination and partial occlusion was studied on different number of images. In these sets of experiments only 3 images were used. For each experiment there were 10 subjects with 11 variations for Yale face database and 11 variation for ORL face database.

The training images used were all with identical properties: with glasses, with no glasses and normal. There are three other images with different lighting condition and 5 images with different facial expressions. These other images were tested against the system trained with the initial three pictures.

Fig 13: A class of image

Besides the images form the data set these images have been occluded with vertical and horizontal distortions. The performance of each algorithm is tested with various distortions in the image.

5.2 Test Results

Five sets of experiments were conducted. Results of the experiments are summarized in the table 5.1 to 5.5. The table shows the performance of each of the algorithms with different corresponding distance metrics. The same training images were used for all the experiments to obtain consistent results with each algorithm.

5.2.1 Using the Normal Yale Face Database

The training images as well as all the other images were tested in this experiment. The experiment was conducted with 30 training images and the total number or images to be recognized were 110 images. The results demonstrating the recognition performance of all four techniques are presented in Table 5.1.

Table 5.1 Results for Normal Yale Face Database

	Euclidean Distance	City Block Distance	Angular Distance	Mahalanobis Distance
Eigenfaces	80%	50.09%	83.6%	72.72%
Fisherfaces	68.18%	61.36%	77.52%	84.09%
ICA1	67.27%	47.27%	85.45%	45.45%
ICA2	81.81%	45.45%	72.72%	63.63%

For this experiment it was found that ICA along with the angular distance metric provided the best recognition measure.

5.2.2 Using the vertically I occluded Yale Face Database

For this experiment all the images in Yale Face Database were occluded with a vertical strip. The strip acts as noise to the face recognition system. The results for the recognition system with this apparatus is presented in Table 5.2. The occluded images are shown in Appendix B

Table 5.2. Results for vertically I occluded Yale Face Database

	Euclidean Distance	City Block Distance	Angular Distance	Mahalanobis Distance
Eigenfaces	85.45%	61.81%	83.63%	63.63%
Fisherfaces	54.54%	50%	72.62%	70.45%
ICA1	34.54%	7.27%	78.18%	4.81%
ICA2	64.54%	18.18%	60.06%	54.45%

For this experiment it was found that PCA along with the Euclidean distance metric provided the best recognition measure.

5.2.3 Using the vertically II occluded Yale Face Database

For this experiment all the images in Yale Face Database were occluded with two vertical strips. The results for the recognition system with these images is presented in Table 5.3.

Table 5.3 Results for vertically II occluded Yale Face Database

	Euclidean Distance	City Block Distance	Angular Distance	Mahalanobis Distance
Eigenfaces	65.45%	21.81%	23.63%	12.72%
Fisherfaces	47.73%	43.18%	60.37%	70.45%
ICA1	14.54%	12.72%	72.72%	27.27%
ICA2	69.1%	16.36%	52.74%	60.6%

For this experiment it was found that ICA along with the Angular distance metric provided the best recognition measure.

5.2.4 Using the Horizontally Occluded Yale Face Database

For this experiment all the images in Yale Face Database were occluded with a horizontal strip. The results for the recognition system is presented in Table 5.4.

Table 5.4: Results for Horizontally Occluded Yale Face Database

	Euclidean Distance	City Block Distance	Angular Distance	Mahalanobis Distance
Eigenfaces	81.81%	56.36%	69.09%	61.81%
Fisherfaces	54.57%	29.45%	57.83%	75%
ICA1	21.81%	5.45%	72.22%	10.9%
ICA2	23.63%	7.27%	52.72%	49.09%

For this experiment it was found that PCA along with the Euclidean distance metric provided the best recognition measure.

5.2.5 Using the Normal ORL Face Database

For the proper comparison of the database on more experiment was conducted with a separate database with different set of images and training images. The ORL database was used in the experiment and the following results were obtained.

Table 5.5: Results for Normal ORL Face Database

	Euclidean Distance	City Block Distance	Angular Distance	Mahalanobis Distance
Eigenfaces	87.2%	70.22%	88.14%	85.5%
Fisherfaces	80.6%	60.5%	79.8%	87%
ICA1	75.95%	42.25%	84.54%	50.28%
ICA2	83.34%	50.18%	79.2%	67.41%

The results showed that the best measure is the Eigenfaces when the distance metric is the angular distance.

5.3 Metric Comparison

Four popular metrics were used for the experiments. Among the four metrics it was found that the Euclidean distance and Angular distance proved to be the most efficient distance metric among all the experiments. In many cases the angular distance metrics even outperformed the popular Euclidean distance The City block distance although being a very frequently used distance metric in the past provided satisfactory results with PCA and LDA only. The Mahalanobis distance provided a good distance measure but its performance could not be used as a satisfactory measure as the results depended on various criterions of the experiments.

Here, none of the metric can be called best and there are no any particular metric to be used with the particular algorithms. At best, it can be stated that it depends on the nature of the task. There are no good combination of algorithm and metric for the purpose of face recognition.

5.4 Comparison to Previous Work

The results showed no clear line of demarcation between the algorithms. The claims made by Bartlett et al. [4] about ICA outperforming PCA and Belhumeur et al [3] about LDA outperforming PCA was not consistent with the results. From the results, the PCA basically showed more potential as a better face recognition system.

The results were more consistent with the claims made by Moghaddam [18] whose claims stated that there was no significant difference between PCA and ICA without any occlusions. The claims made by Martinez [5] were also accurate as it was found that LDA necessarily does not outperform PCA but depended on various criterions.

CHAPTER 6: CONCLUSIONS

6.1 Conclusion

The thesis was focused on independent, comparative study of three most popular appearance based face recognition algorithms (PCA, ICA and LDA) in completely equal working condition with different metric combination. It was found that no algorithm metric combination is perfect at the moment. The comparison and implementation of these algorithms for more accurate results require further research. There must be a deeper understanding of each algorithm for the proper implementation which demands the use of the strengths of the algorithm and removal of the weakness which hinder the recognition purposes.

From the results it was concluded that the PCA despite being the first of the appearance based algorithm is still a very strong algorithm for the purpose of Face Recognition. One might assume that the separation of classes in LDA and separation of individual properties in ICA would result in much better recognition results, but it is not so. The human face is a very complex piece of information which still needs research for proper understanding and proper representation.

In the results, it was seen that the performance of ORL face database outperformed the Yale face database. This was due to the faces that the ORL face database had been properly cropped so that there was no information on the image except the face. The Yale face database was found as a good representation with different variation in lighting and emotions but it needed more preprocessing steps and could not be used directly for recognition purposes as compare to ORL face database.

6.2 Further Studies

In this thesis, the algorithms PCA, LDA and ICA were studied. The face recognition technology is evolving even more with many more feature based algorithms and appearance based algorithms. Although the appearance base algorithms are popular now; the performance of the feature based techniques cannot be denied.

The algorithms studied on different disciplines were found as a good measure and a lot of efforts have been put to their research, but they still have shortcomings. The PCA, LDA and ICA studied here showed a good recognition rate but they were not the best. There were different conditions where each algorithm worked and the other didn’t. The shortcomings of the algorithms have been addressed in the hybrid face recognition system.

REFERENCES

[1] L. Sirovich and M. Kirby, "Low dimensional procedure for the characterization of human faces", Journal of Optical Society of America A, Vol 4 Pg519, March 1987

[2] Turk M. and Pentland A., "Face recognition using eigenfaces", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, 1991

[3] Belhumeur N., Hespanha J. and Kriegman D., "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", Proc. ECCV, 45--58, 1996

[4] M. Bartlett, J. Movellan, and T. Sejnowski, "Face Recognition by Independent Component Analysis" IEEE transaction on Neural Networks, Vol. 13, No. 6, November 2002

[5] Aleix M. Martinez, Avinash C. Kak., "PCA vs LDA", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2 February 2001

[6] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995

[7] V. C. Chen, "Evaluation of Bayes, ICA, PCA and SVM Methods for Classification", Radar Division, US Naval Research Laborator

[8] W. Zhao, R Chellappa, A. Rosenfled, P.J. Phillips "Face Recognition: A literature Survey"

[9] Gregory Shakhnarovich, Baback Moghaddam, "Face Recognition in Subspaces", Mitsubishi Electric Research Laboratories

[10] Luis Torres, "IS THERE ANY HOPE FOR FACE RECOGNITION?", Technical University of Catalonia, Barcelona, Spain

[11] James V. Stone, "Independent Component Analysis", Encyclopedia of Statistics in Behavioral Science Volume 2, pp. 907–912

[12] A. Hyvärinen and E. Oja, “Independent Component Analysis: Algorithms and Applications”, Neural Networks Research Centre, Helsinki University of Technology

[13] P. Comon., “Independent component analysis, a new concept?”, Signal Processing, 36:287–314, 1994.

[14] R. Chellappa, C. Wilson and S. Sirohey, “Human and Machine Recognition of Faces: A Survey”

[15] W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng, “Discriminant Analysis of Principal Components for Face Recognition”, pp. 336-341, International Conference on Automatic Face and Gesture Recognition, 1998.

[16] J.-F. Cardoso, "Infomax and Maximum Likelihood for Source Separation," IEEE Letters on Signal Processing, vol. 4, pp. 112-114, 1997.

[17] K. Baek, B. Draper, J.R. Beveridge, K. She, "PCA vs. ICA: A Comparison on the FERET Data Set", Proc. of the Fourth International Conference on Computer Vision, Pattern Recognition and Image Processing, Durham, NC, USA, 8-14 March 2002, pp. 824-827

[18] B. Moghaddam, "Principal Manifolds and Probabilistic Subspaces for Visual Recognition", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 6, October 2002, pp. 780-788

[19] V. Bruce, “Identification of Human Faces”, pp. 615-619, Image Processing and Its Applications, Conference Publication No. 465, IEEE, 1999

[20] G. Shakhnarovich and B. Moghaddam, “Face Recognition in Subspaces”, ITSUBISHI ELECTRIC RESEARCH LABORATORIES

[21] B. Draper, K. Baek, M. Bartlett and J. Beveridge, “Recognizing Faces with PCA and ICA”, Computer Vision and Image Understanding 91 (2003) 115–137

[22] Naotoshi Seo, “Eigenfaces and Fisherfaces”, University of Maryland, ENEE633 Pattern Recognition, Project 2-1

[23] Lieven De Lathauwer, Bart De Moor, Joos Vandewalle, “An introduction to independent component analysis”, Journal of Chemometrics, J. Chemometrics 2000; 14: 123–149

[24] M. Zibulevsky and B. A. Pearlmutter, "Blind Separation of Sources with Sparse Representation in a Given Signal Dictionary," presented at International Workshop on Independent Component Analysis and Blind Source Separation, Helsinki, 2000.

[25] Chengjun Liu and Harry Wechsler, "Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition" *Second International Conference on Audio- and Video-based Biometric Person Authentication*, AVBPA'99,Washington D. C. USA, March 22-24, 1999

[26] Leonid Zhukov, David Gleich, "Topic Identification in Soft Clustering using PCA and ICA"

[27] Rao, S.B, Shantha, C.K., "Numerical Methods", Universities Press, Chapter 8

[28] Gupta, S. K., "Numerical Methods for Engineers", New Age International (P) Limited Publishers, pp 47-55

[29] Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., "Numerical recipes in C", Second Edition, Cambridge University Press (Indian Reprint), pp. 469-481

[30] Robert P.W. Duin and Elzbieta Pekalska, "Open issues in pattern recognition"

[31] J. P. Marques de Sa, Pattern Recognition Concepts, Methods and Applications

[32] Sing Tze Bow, Pattern Recognition and Image Processing

[33] http://www.unc.edu/~smithdr/inls187/biofuture.htm

[34] http://en.wikipedia.com

[35] http://people.revoledu.com/kardi/

[36] http://cvc.yale.edu/projects/yalefaces/yalefaces.html

[37] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

Appendix A

Reference for the symbols

Symbol	Meaning
x	Width of image
y	Height of image
M	The number of sample images
k	The number of eigenfaces to be generated. The value can be altered to vary the performance of the system; however $k \leq M$
t	The number of individuals known to the classification system. As a rule, $t \leq M$ and $k << t$
$\tilde{A}_1 \ldots \tilde{A}_M$	These are the sample images as column vectors. Each sample image of uniform size x pixels across and y pixels down. The size of the vectors is x*y
Ø	This is the average image found from the sample images. This column vector is of the same size as the sample images vectors
$Øc_{1 \ldots M}$	This is the average image of a particular class c
$\ddot{O}_1 \ldots \ddot{O}_M$	These vectors are the difference between each sample image and the average image. These column vectors are of the same size as the sample images vectors
A	This is the matrix generated by considering each of the Ö vectors as the column of this matrix. The dimension of this matrix is (x*y)*M. where M is the number of sample images
$\lambda 1 \ldots \lambda_k$	These are the eigenvalues of the A^TA and AA^T matrices
$X_1 \ldots X_k$	These are the eigenvectors for the corresponding eigenvalues
$u_1 \ldots u_k$	These vectors are the eigenfaces generated from the sample images
$\mathbf{U}=[u_1 \ldots u_k]$	Matrix of eigenfaces
$\mathbf{\grave{U}}=\{\grave{u}_1,\grave{u}_2,\ldots,\grave{u}_k\}$	**Ù** represents the vector of weights that result from a projection into facespace. Each *ù* is a scalar such that when each is multiplied by its corresponding eigenface and then all the weighted eigenfaces are added together, the original image results.
S_W	Within Class Scatter Matrix
S_B	Between Class Scatter Matrix
W	Weights obtained from LDA

For Chapter 3.2 and 3.3