Winter 2009

A Subspace Projection Methodology for Nonlinear Manifold Based Face Recognition

Praveen Sankaran
Old Dominion University

Recommended Citation

Sankaran, Praveen. "A Subspace Projection Methodology for Nonlinear Manifold Based Face Recognition" (2009). Doctor of Philosophy (PhD), Dissertation, Electrical & Computer Engineering, Old Dominion University, DOI: 10.25777/5v3e-vc49
https://digitalcommons.odu.edu/ece_etds/121

A SUBSPACE PROJECTION METHODOLOGY

FOR NONLINEAR MANIFOLD BASED FACE RECOGNITION

by

Praveen Sankaran
B.Tech., August 2002, University of Calicut, India
M.S. August 2005, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY
December 2009

Approved by:

Vijayan K. Asari (Director)

Mohammad A. Karim (Member)

Zia-ur Rahman (Member)

Jiang Li (Member)

Norou Diawara (Member)

ABSTRACT

A SUBSPACE PROJECTION METHODOLOGY FOR NONLINEAR MANIFOLD BASED FACE RECOGNITION

Praveen Sankaran
Old Dominion University
Director: Dr. K. Vijayan Asari

A novel feature extraction method that utilizes nonlinear mapping from the original data space to the feature space is presented in this dissertation. Feature extraction methods aim to find compact representations of data that are easy to classify. Measurements with similar values are grouped to same category, while those with differing values are deemed to be of separate categories. For most practical systems, the meaningful features of a pattern class lie in a low dimensional nonlinear constraint region (manifold) within the high dimensional data space. A learning algorithm to model this nonlinear region and to project patterns to this feature space is developed. Least squares estimation approach that utilizes interdependency between points in training patterns is used to form the nonlinear region. The proposed feature extraction strategy is employed to improve face recognition accuracy under varying illumination conditions and facial expressions. Though the face features show variations under these conditions, the features of one individual tend to cluster together and can be considered as a neighborhood. Low dimensional representations of face patterns in the feature space may lie in a nonlinear constraint region, which when modeled leads to efficient pattern classification. A feature space encompassing multiple pattern classes

can be trained by modeling a separate constraint region for each pattern class and obtaining a mean constraint region by averaging all the individual regions. Unlike most other nonlinear techniques, the proposed method provides an easy intuitive way to place new points onto a nonlinear region in the feature space. The proposed feature extraction and classification method results in improved accuracy when compared to the classical linear representations.

Face recognition accuracy is further improved by introducing the concepts of modularity, discriminant analysis and phase congruency into the proposed method. In the modular approach, feature components are extracted from different sub-modules of the images and concatenated to make a single vector to represent a face region. By doing this we are able to extract features that are more representative of the local features of the face. When projected onto an arbitrary line, samples from well formed clusters could produce a confused mixture of samples from all the classes leading to poor recognition. Discriminant analysis aims to find an optimal line orientation for which the data classes are well separated. Experiments performed on various databases to evaluate the performance of the proposed face recognition technique have shown improvement in recognition accuracy, especially under varying illumination conditions and facial expressions. This shows that the integration of multiple subspaces, each representing a part of a higher order nonlinear function, could represent a pattern with variability. Research work is progressing to investigate the effectiveness of subspace projection methodology for building manifolds with other nonlinear functions and to identify the optimum nonlinear function from an object classification perspective.

This dissertation is dedicated to my family and friends.

ACKNOWLEDGMENTS

I was immensely supported by a group of people during my entire graduate school. I would first like to thank my family for the patience they have shown for the last 6 years. The main force behind my Ph.D. is my advisor Dr. Vijayan Asari, whose support and advice have helped me overcome many a difficult situation. I thank Dr. Zia-ur Rahman, Dr. Mohammad Karim, Dr. Jiang Li, Dr. Nazrul Islam, and Dr. Norou Diawara for the time and energy they dedicated to me, giving suggestions and advice on my research work.

Acknowledgments are due to my friends at ODU Vision Lab for their support and cooperation during my research work. My early work was greatly influenced by the guidance and advice of Rajkiran, Dr. Ming Seow and Dr. Satyanadh. Special thanks are due to Dr. Hau Ngo for being a great elder brother. The latter part of my work is heavily influenced by the interactions I have had with Jacob Foytik. The stunts and pranks pulled together with Cort Tompkins, Loc Tran and the "new kids", have resulted in many lighter moments. I believe the friendship I have been able to form with them has had a very positive influence on my outlook towards life.

I would like to express my heartfelt thanks to Dr. Sacharia Albin for all the curricular advice and recommendations. I would like to thank Linda Marshall and Romina Samson for helping me with all the confusing paper work and "procedures" of graduate school.

TABLE OF CONTENTS

Section Page

Section Page

LIST OF FIGURES

LIST OF TABLES

Table Page

1 INTRODUCTION

The process of learning involves such tasks as associating a name with an object or understanding speech or writing through examples. This is similar to the idea of training a machine with sample data. The process of recollection involves assigning the input data as one of the trained categories. Duda, Hart and Stork [1] define the process of pattern recognition as the act of taking in raw data and making an action based on the "category" of the pattern.

Any pattern classification system consists of at least three main blocks as shown in Figure 1.1. A sensor converts physical input into processable data. In a human system various sensory organs, such as eyes for sight, perform this role. Similar to this is the role of a camera in case of a machine. A good sensor is important for later steps. For example, classification accuracy of an image would depend on the quality of image obtained from the camera. The initial sensing step is usually followed by some sort of pre-processing of data for aiding in better feature extraction.

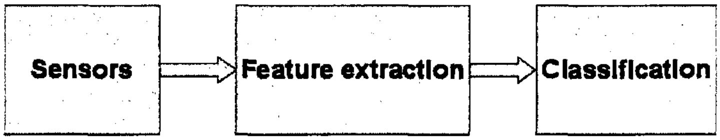


Figure 1.1: Pattern classification system.

The role of the feature extractor is to identify distinguishable values for each category. These identified values then represent each object as its featuies. These features tend to be similar for objects in the same category and very different for objects of different categories. In an ideal scenario, features are selected such that

This work is written as per the *IEEE Transactions* format they are invariant to random noises in the input. The classifier looks at the feature values and tries to determine a category for the data under consideration. Some of the applications of enabling a machine to recognize patterns are listed in Table 1.1 [2].

Table 1.1: Examples of pattern recognition applications.

Problem Domain	Application	Input Pattern	Pattern Classes
Bio-informatics	Sequence analysis	DNA/Protein sequence	Known types of genes/patterns
Data mining	Searching for meaningful patterns	Points in multi-dimensional space	Compact and well separated clusters
Document classification	Internet search	Text document	Semantic categories
Document image analysis	Reading machine for the blind	Document image	Alphanumeric characters, words
Industrial automation	Printed circuit board inspection	Intensity or range image	Defective/ non-defective nature of products
Multimedia database retrieval	Internet search	Video clip	Video genres
Biometric recognition	Personal identification	Face, iris, fingerprint	Authorized users for access control
Remote sensing	Forecasting crop yield	Multi-spectral image	Land use categories, growth pattern of crops
Speech recognition	Telephone directory inquiry without operator assistance	Speech waveform	Spoken words

Feature selectors also usually play the role of reducing the dimension of data under consideration. Intuition tells us that a higher number of features should lead to higher accuracy. This is true as long as the features are completely independent of each other. In practice the accuracy levels can fall off with an increasing number of features [1]. High dimensional data also have problems with requirement of a large sample base and higher computational expense.

The simplest way to reduce the number of dimensions and choose features is to use a linear combination of input features. Methods like the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) follow this strategy. These methods project data onto a unit vector defining each dimension. The direction of this unit vector is represented by a straight line, thus forming a linear constraint region in the lower dimension within the high dimensional data space, but images of similar visual perception reside in a nonlinear contraint surface in the low dimensional image space [3]. In the proposed work, this nonlinear region is obtained by modeling the direction of the unit vector as a second order polynomial curve. Thus three different subspaces, each corresponding to the order of the polynomial equation, contribute to the final shape of the constraint region. A learning algorithm to model this nonlinear region based on least squares estimation approach that utilizes interdependency between points in training patterns is proposed in this dissertation.

1.1 Focus and contributions

The main focus of this research is to develop a nonlinear subspace projection method for projecting raw data points to feature space. From the list of problems given in Table 1.1, we consider the problem of biometric recognition with face images as the input data. Specific objectives of this dissertation research are as follows:

1. Development of a nonlinear subspace representation for a data set based on a second order polynomial curve, achieving a nonlinear combination of principal components from different orders to obtain required optimal features.

2. Development of a face recognition system based on the above approach for extracting facial feature components.

3. Improvement of face recognition accuracy by following a discriminant approach to obtain distinguishable clusters in the nonlinear subspace.

4. Application of a modular technique to better model local data using multiple nonlinear subspaces.

5. Reconstruction of data from the feature space.

6. Application of the phase congruent technique to extract features which could overcome illumination variations.

7. Testing and validation of the above techniques on various face databases.

1.1.1 Nonlinear subspace projection

A linear combination of principal components fails to model the underlying nonlinearity of raw data in feature space. Many methods that do model this nonlinearity fail to present a simple system to project new test points to the feature space making them difficult to work with. These methods also fail to provide a way to recover original information from the transformed data. We propose here a system where a nonlinear combination of the principal components is used to project data onto the feature space as illustrated in Figure 1.2. We chose to model the data as a second order polynomial curve, which allows us to show the validity of combining subspaces to obtain the desired nonlinearity. Components in each of these subspaces is modeled using a least squares estimation approach on the raw data. The three separate subspaces are then combined using the second order polynomial equation.

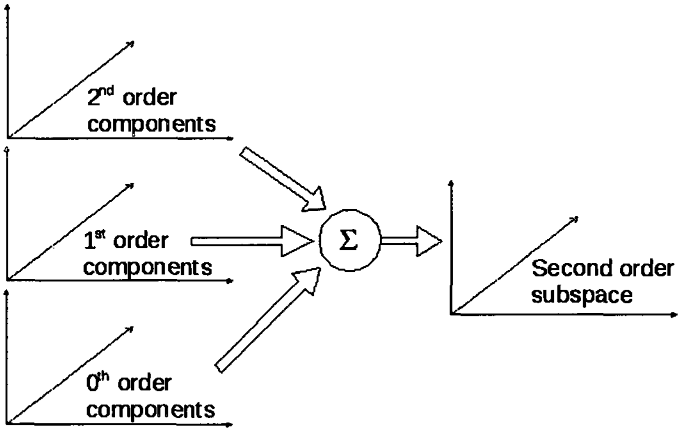


Figure 1.2: Nonlinear subspace combination.

1.1.2 Face recognition system

The face recognition task involves identifying faces in a still or video sequence using a stored database. There has been tremendous development in this area with the development of a number of methods to match one face to another. The various methods for face recognition mainly fall under two types – appearance based and geometry based. Appearance based methods depend on space transformation techniques like PCA, ICA (Independent Component Analysis) or LDA to code the faces, thus, in effect creating a template for each face in a database. The input test image is transformed to the same space model and compared to the templates in the database to find the closest match as shown in Figure 1.3. Subjects 2 and 3 are easily classified in this case, but there is ambiguity in the case of subject 1, and such cases could lead to incorrect classification.

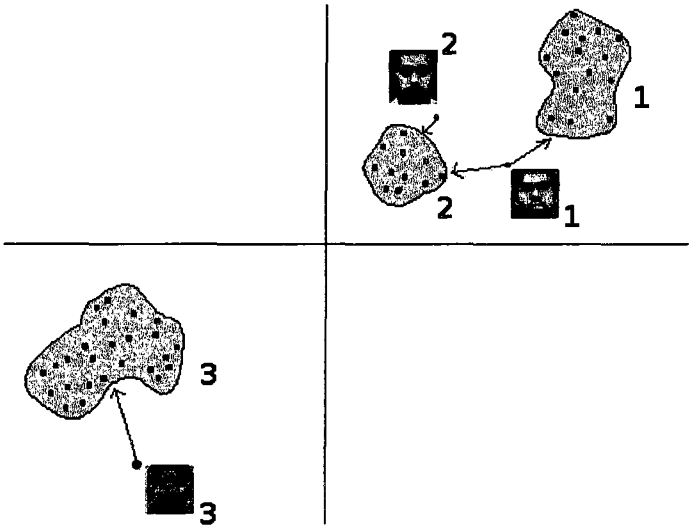


Figure 1.3: Face clusters.

An outline of a face recognition system is shown in Figure 1.4. Face images stored in a database are transformed to the useful feature space, and only these transformed values are retained after the feature extraction process. A test input undergoes transformation to the same subspace as the trained set and a feature classifier determines the category or the identity of the input face.

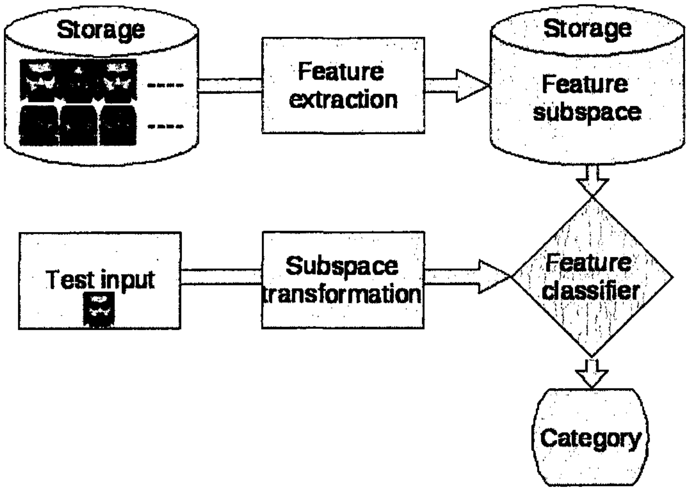


Figure 1.4: Face recognition flow.

1.1.3 Discriminant approach

Samples from well formed clusters, when projected onto an arbitrary plane could produce a confused mixture of samples from all the classes leading to poor classification. Discriminant analysis aims to find optimal plane orientations for which the data classes are well separated. Figure 1.5 shows the difference between PCA and LDA by projecting 2D points from two clusters. When projected onto the line formed using PCA (shown in Figure 1.5a), the clusters can result in a mixture that is now hard to classify. LDA projection (shown in Figure 1.5b) results in two easily separable clusters.

1.1.4 Modular technique

In modular technique, an image is divided into R sub-images, where R is a constant, and weight vectors are found separately for each of the sub-images. In a global approach, we find one set of directions onto which we project all the face features,

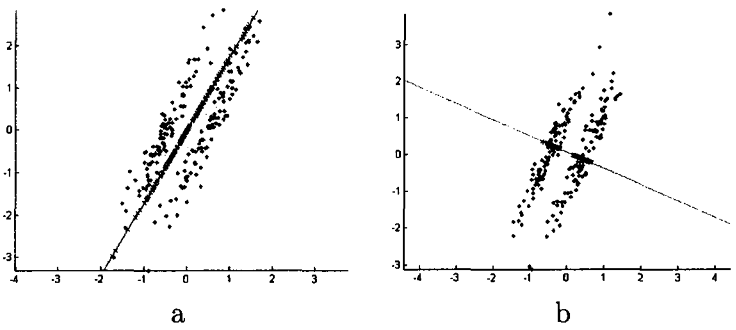


Figure 1.5: Data cluster on first two principal components. a. PCA, b. LDA.

but given the variability of different regions of a face image, it makes much better sense to divide a face image into different regions and model subspace projection directions for each of them separately. Concatenation of the weight vectors of all the different sub-images forms a signature weight vector. The idea can be looked at from the point-of-view of using multiple simple directional curves to model one complex curve. Thus, we have an additional layer of nonlinearity in feature extraction when using the modular approach.

1.1.5 Phase congruent features

One of the key issues in face recognition is the image distortion due to varying illumination conditions. Recognition accuracy could be improved by using features that are invariant to illumination. Phase congruent features are the locations in an image where the frequency components are maximally in phase and these features are invariant to changes in image brightness and contrast. The proposed modular nonlinear subspace projection technique is now applied to phase congruent images.

1.2 Thesis outline

This dissertation is presented in the following order. Section 2 provides an overview of nonlinear methods for feature extraction followed by a detailed report on linear feature extraction strategies. We concentrate on applying feature extraction methods for face recognition. This section also looks into issues faced in face recognition and how linear methods are not ideal when faced with those problems. Section 3 discusses nonlinear methods for feature extraction, with an emphasis on one key method and its extension for pattern classification. Section 4 gives a detailed discussion of the development of the proposed method. Section 5 details four different standard test databases used for the purpose of testing the algorithms along with simulation results and discussions. Section 6 provides the conclusion and direction for future work.

2 Literature Review

A review of linear and nonlinear approaches for pattern classification is presented in this section. The review focuses on component analysis methods especially related to face recognition and their weaknesses and details a brief overview of nonlinear manifold methods. A detailed review of the nonlinear methods will be given in Section 3.

High dimensional (raw) data is difficult to understand and analyze. The dimension of data is the number of variables that are measured on each observation. For example, an image can be considered as an observation. In this case, each pixel value inside the image is a variable. Thus, the number of pixels under consideration becomes the number of dimensions we are dealing with. In most cases not all these variables are needed to describe the observation efficiently. Having a large number of variables also requires a large number of observations to classify these variables (curse of dimensionality [1]). Dimensionality reduction methods aim to find compact representations of data. Methods that try to solve this problem can be divided into linear and nonlinear classes. Linear methods include PCA [4], ICA [5], etc. PCA is commonly used to perform dimensionality reduction by projecting data into a subspace spanned by the eigenvectors of the covariance matrix of the data. In computer vision applications, the PCA method has been used for the representation and recognition of faces [6], image super-resolution [7], hand-print recognition [8] and facial feature extraction [9]. PCA is simple but efficient only for data having low dimensional linear structure. This approach will not work when dealing with manifolds of high curvature as nonlinear structures in data sets are approximated to a linear structure in PCA.

2.1 Overview of nonlinear methods

A set of face images scattered randomly while represented in high dimensional vector form, is known to be constrained to a specific format in a much lower number of dimensions. The meaningful features in a face data set have been shown to lie in a low dimensional nonlinear manifold or constraint surface [10]. Tenenbaum et al. also presented a method to model the manifold based on a number of nearest neighbors and the geometric distance between the points in a data set (isometric feature mapping, known as Isomap). A comparable method is the Locally Linear Embedding [11] that tries to preserve the geometric properties of the data in the projected low dimensional space. These nonlinear methods generally unfold nonlinear low dimensional manifolds that cannot be modeled by linear methods. The main steps in their implementation can be summarized in the following three steps.

1. Compute distances in input space and defining neighborhoods.
2. Based on the defined neighborhood, compute a square matrix, each row representing an element in the input space.
3. Low dimensional embedding of the data using eigenvectors of the above matrix.

The issue here, though, is that these methods require us to compute an immediate neighborhood of any new point using Euclidean distances to get paths and geometric properties. So with the introduction of a new point, it is required to recompute the neighborhoods of an entire data set as the new point could be a neighbor of one or several points already in the database. This batch mode of computation is not practical when dealing with video streams of faces and when real time recognition is required. Incremental modifications proposed in [12] would still require us to modify the neighborhood graphs before projecting the new point. Another issue with these methods has been the way the neighborhood is defined. There are two general ways we define the neighborhood, through an integer k that specifies the number of neighbors or through a distance threshold ε. It's been found that optimal values of either of these parameters vary for each database we consider and that final accuracy values depend greatly on the optimality of these two parameters [13]. While the concept of nonlinearity is important, we need to come up with a method that can model the low dimensional nonlinearity and can provide a direct projection to the nonlinear low dimensional space.

Seow et al. [14, 15] modeled the manifold as an associative memory using a recurring neural network and presented the concept of representing data as lines instead of point representation. The idea of varying face images (of the same person) due to varying pose, illumination conditions or expression belonging to a curve is presented in [16]. Instead of presenting a face image as a single vector in the projected space, a face image is now represented using the coefficients of a multi-ordered curve. A collection of these curves models the underlying nonlinearities in the database. The approach proposed in this dissertation finds its base in this idea of representing data patterns.

2.2 Face recognition

The face recognition task involves identifying faces in a still or video sequence using a stored database. The task is complex due to various factors such as varying illumination conditions, poses, expressions and variations due to aging. An ideal face recognition system should be insensitive to all these variations. Research on automatic face recognition has been pursued by several scientists all over the world for over three decades. Chellappa et al. [17] presented a survey on several methods for face recognition. Early attempts to recognize faces were conducted in the mid-1970's using pattern classification methods. These methods used face features from faces or face profiles to do the task. The survey also lists the contributions by psycho-physicists and neuroscientists. They have been concerned with issues such as uniqueness of faces, analysis of facial expressions for face recognition, how infants perceive faces, organization of memory for faces, etc.

Face recognition methods mainly fall under model based or appearance based algorithms. Various geometrical parameters are considered in the model-based approach given in [18]. A set of geometrical features, such as nose width and length, mouth position, and chin shape are computed. A comparative study of this method with a template based method presented in [18] reported better recognition accuracy for the template method. Appearance based methods depend on spacial transform techniques like PCA, ICA or LDA to code faces, thus creating a template for each face in a database. The input test image is transformed to the same space model and compared to the templates in the database to find the closest match. The first attempt at representing faces using principal components was by Sirovich and Kirby in 1987 [19]. This paper demonstrated that any face can be economically represented in terms of a best coordinate system called the eigenpictures. A face with 2^{14} pixels could be characterized by 40 numbers to within a 3% error. This statistical approach expressed face images as a subset of their eigenvectors and are therefore called eigenfaces. Figure 2.1 shows typical eigenfaces of the original image shown in the top-left corner [20].

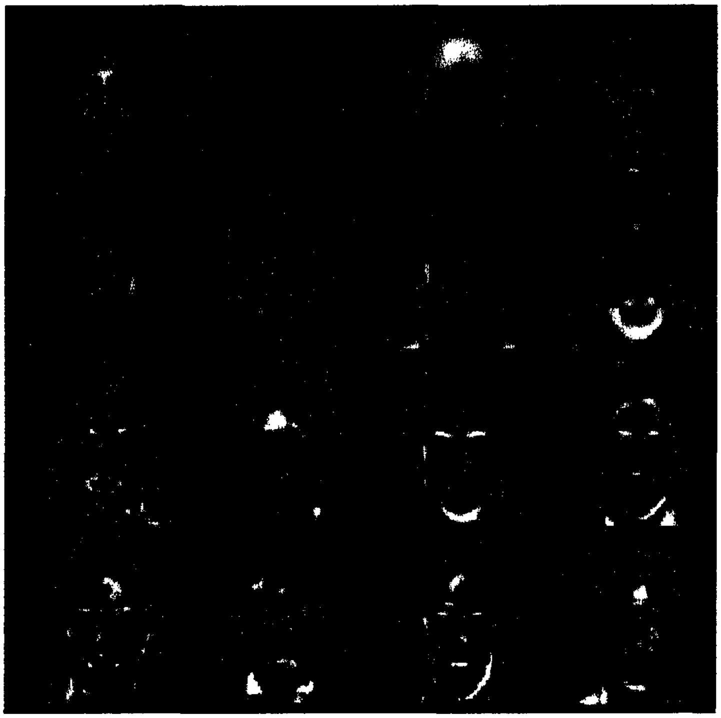


Figure 2.1: Standard eigenfaces.

The PCA method was used for face recognition in 1991 by Turk and Pentland [6]. The attempt was to build real time face detection and recognition system based on eigenfaces. This research laid down a specific process for recognition as given below:

1. Acquire the training set of face images. Calculate the eigenface that de-

fines the face space.

2. Calculate the weights for the test image based on a certain number of eigenfaces by projecting the input image onto each of the eigenfaces.

3. Check to see if the input image is close enough to the face space to be regarded as a face image.

4. If it is a face, classify the pattern as a known or unknown face.

5. If the same unknown face is seen many times over, try to train it.

Figure 2.2 illustrates a systematic approach to a face recognition system for general surveillance. The input image can be a still image or a frame taken from a video stream. Face detection methods try to find regions in the image that resemble a human face. Feature localization ranges from estimating eye and mouth edges to locating eye centers. There are about 150 reported face detection techniques presented in the literature for both gray-scale and color images [21]. Geometrical approaches based on facial features such as eyes, nose and mouth were implemented by Gee and Cipolla [22] and Horprasert et al. [23]. Various studies in this field also include the Radial Basis Function network estimator by Beymer et al. [24] and the neural network based system by Rowley et al. [25], which are capable of detecting faces with rotation in the image plane. OpenCV face detector based on the Viola-Jones method is used here for face detection [26]. Eye centers are used to normalize image positions, and this is found to help the recognition process.

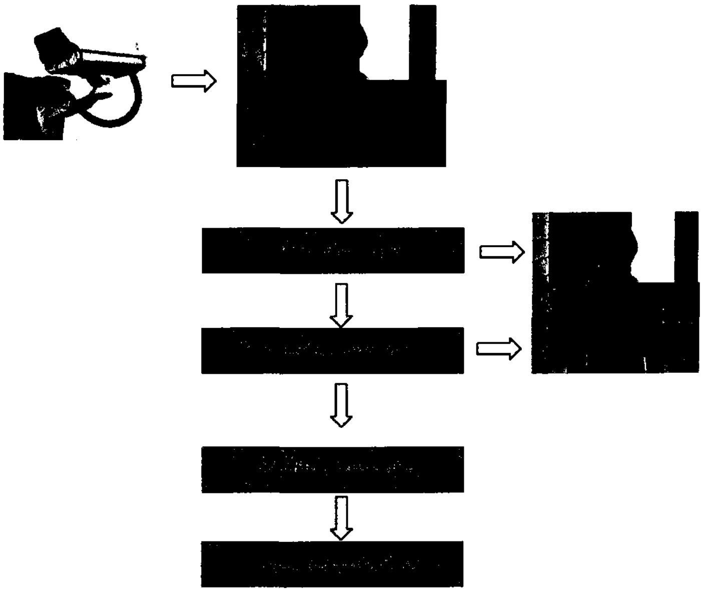


Figure 2.2: Block diagram representation of face recognition system.

2.3 Face recognition with PCA

PCA aims to encode the relevant information of a face image as efficiently as possible. The variation information in a collection of face images is used to encode and compare individual face images. The principal components or the eigenvectors of the covariance matrix of the set of face images are found. Each image can be represented as a linear combination of the eigenvectors. Here we can approximate by using only the best possible eigenvectors, those corresponding to the largest eigenvalues.

2.3.1 Training

Consider the face images to be of size $N \times N$. These images are represented by a vector of size N^2. Since the face images have a similar structure, the vectors representing them will be correlated. These vectors define the subspace of the face images which is called the "face space". Due to the correlation, the images can be represented by a lower dimensional space. Let $I_1, I_2, \ldots I_M$ be the training set of face images. The average image is found by,

$$Y = \frac{1}{M}\sum_{i=1}^{M} I_i \tag{2.1}$$

The vector $A_i = I_i - Y$ is the difference image of each face image. The covariance matrix is obtained from the difference vectors as,

$$C = \frac{1}{M}\sum_{i=1}^{M} A_i.A_i^T \tag{2.2}$$

The eigenvectors of the covariance matrix are computed, and those corresponding to the largest $d\,(d \lll N)$ eigenvalues are selected. The weight vector for each image is computed using these eigenvectors,

$$W_{ik} = E_k^T.\,(I_i - Y) \quad \forall i, k \tag{2.3}$$

where, E_k are the eigenvectors corresponding to the d largest eigenvalues of C, and k varies from $1\,to\,d.$

2.3.2 Testing

The test image is transformed to weight vector form using,

$$W_{testk} = E_k^T.\,(I_{test} - Y) \quad \forall i, k \tag{2.4}$$

The mean weight vector is used to fit the test image to a predefined test class. A simple method to do this would be to find the minimum distance from the test weight vector W_{testk} to the mean weight vector T_p. The nearest neighbor is found using,

$$D_k = \|W_{testk} - W_{ik}\| \quad \forall i, k \tag{2.5}$$

2.3.3 Probabilistic classification

The simplest way to form a similarity measure between two images I_1 and I_2 is the norm $\|I_1 - I_2\|$. The similarity measure $S(I_1, I_2)$ can be set inversely proportional to the norm [27]. This method gives equal weight to all variations, thus ignoring the fact that some variations are more critical than the rest and need to be given more importance. A probabilistic similarity measure, based on the probability that the image intensity difference $\Delta = I_1 - I_2$ denotes variation in appearance of same pattern, can be formulated. For the purpose of face recognition, two different types of variation are defined, intra-personal variations Ω_I and extra-personal variations Ω_E. A similarity measure based on the two types of variations using Bayes rule can be expressed as,

$$S(\Delta) = P(\Omega_I \mid \Delta) = \frac{P(\Delta \mid \Omega_I) P(\Omega_I)}{P(\Delta \mid \Omega_I) P(\Omega_I) + P(\Delta \mid \Omega_E) P(\Omega_E)} \tag{2.6}$$

$P(\Delta \mid \Omega_I)$ and $P(\Delta \mid \Omega_E)$ can be calculated from the training data and values for two $P(\Omega)$ depend on some knowledge of the data under consideration. Two images are considered to belong to the same class if $S(\Delta) > \frac{1}{2}$.

The parameter Δ has the same dimensions as the image vectors from which it was computed. PCA is used to obtain a principal surface whose principal components are used to form the low dimensional estimate using only the first d components $b = [b_1 \ldots b_d]$. Each image can now be stored using weights given by,

$$b_{\Omega_I} = \Lambda_I^{-\frac{1}{2}} V_I I \tag{2.7}$$

where Λ and V are the matrices of the largest eigenvalues and eigenvectors of $\sum_I$ (covariance matrix of intra-personal variations).

2.4 Metric Multi-dimensional Scaling (MDS)

MDS [28] transforms a distance matrix to a set of coordinates. The distances recovered from these coordinates approximates the original distance. Let X be an N dimensional input data matrix with M points $[x_1 \ldots, x_M]$. A between observation cross product matrix can be formed by,

$$S = X \times X^T \tag{2.8}$$

A squared distance matrix can be formed from the above cross product matrix as,

$$D = S\alpha^T + \alpha S^T - 2S \text{ where} \, \alpha \rightarrow [1_1 \cdots 1_M] \tag{2.9}$$

The idea behind MDS is to convert Equation 2.9 so that we have a cross product matrix from the distance matrix. For this a centering matrix is defined as,

$$\Omega = I - \alpha^T \tag{2.10}$$

Then, the cross product matrix S can be obtained from the distance matrix D as,

$$S = -\frac{1}{2}\Omega D \Omega^T \tag{2.11}$$

The eigenvectors of the above S matrix provides the projection factors required.

2.5 Modular approach

The classical PCA technique is not effective for face recognition in varying pose and illumination conditions. The weight vectors vary considerably from the weight vectors with frontal pose at normal illumination; hence, identification becomes difficult. The modular PCA (mPCA) [29, 30] is an approach proposed to overcome the low accuracy of PCA in cases of varying illumination, facial expressions and pose. In mPCA the image is divided into R disjoint blocks and an average sub-space from all the different blocks is considered. Weight vectors are formed for each image module. The method showed improvement in cases of illumination variation but did not produce better results in cases of varying pose.

2.6 Issues in face recognition

The major problem for face recognition applications is variations in face images, resulting in corresponding large variations in the feature space within a class due to:

1. Variations in illumination,

2. Differing facial expressions,

3. Varying face poses,

4. Occlusion of regions of face due to facial hair, glasses, etc.,

5. Natural variations in face due to aging or injuries.

Figures 2.3 to 2.5 show images with varying illumination, poses and expressions. While a human can make out that the sequence is of the same person within each variation, for a machine pixel values change drastically across each change leading to possible errors in classification.

Figure 2.4: Faces with varying poses

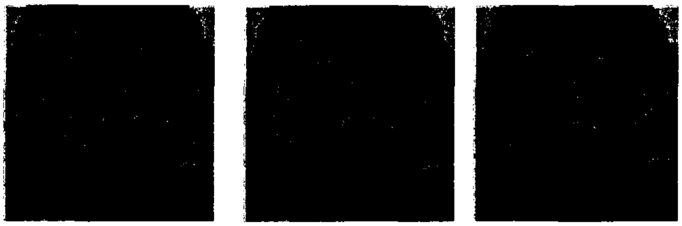


Figure 2.5: Faces with varying expressions.

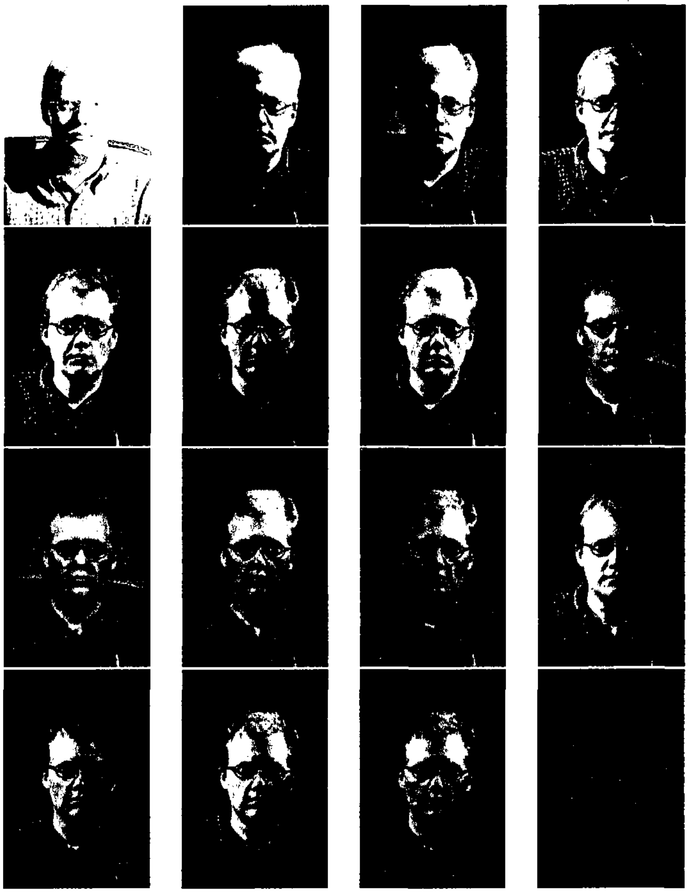


Figure 2.3: Faces with varying illumination.

2.7 Summary

The classical techniques work well in controlled environments but fail in complex environments with the problems listed in section 2.6. One of the goals of this research is to improve the performance of face recognition systems over that of classical methods. Various techniques have been proposed to extend the classical methods to improve their recognition accuracy. The modular-PCA [29, 30] method is one such technique that improved the performance of classical PCA based face recognition especially in cases with varying illumination. This approach, however, could not solve issues with pose variations. Varying illumination, pose and facial expressions induces nonlinearity in the face image set placing these points away from the rest of the cluster. The linear methods fail to model this nonlinearity, thus projecting these points to an incorrect cluster in the low dimensional space. This leads to considerable research in methods to model nonlinearities in data to improve classification accuracy. Some of these methods will be discussed in Section 3.

3 Nonlinear Feature Spaces

The aim of dimensionality reduction methods is to find meaningful low dimensional structures hidden in high dimensional structures of original observation data. Linear methods like PCA are simple to implement, computationally efficient and discover true low dimensional structures on or near a linear sub-space of the high dimensional data. The issue with using the linear methods for representing faces is illustrated in Figure 3.1 [10]. The images are of one face observed under different pose and lighting conditions. The original dimensionality of the images is 64 × 64 leading to a 4096-dimensional input space. Within this large dimensional space, all the images can be represented as a 3 dimensional manifold defined by two pose angles and a lighting direction. Some of the methods that attempt to model this nonlinearity are discussed in this section.

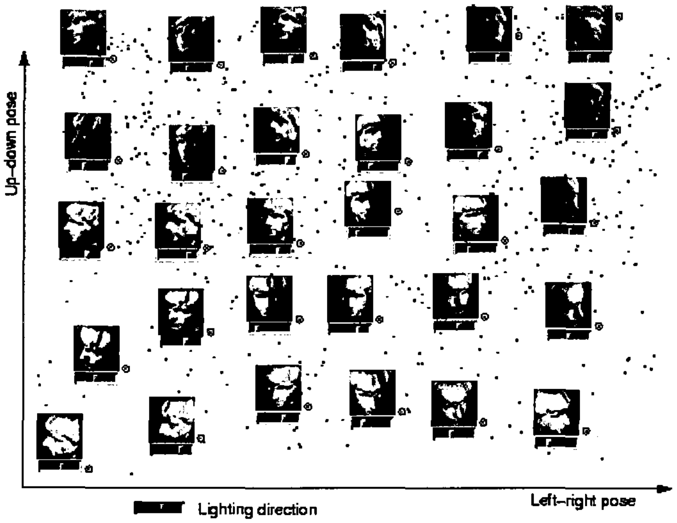


Figure 3.1: Synthetic face set.

3.1 Kernel methods

Kernel methods try to solve the problem of nonlinearity by transformation from image space to a feature space that has a higher dimension. The two data clusters, as shown in Figure 3.2a are not classifiable by a linear classifier. The same two clusters when projected to a feature space with higher dimension, as in Figure 3.2b, are easily separated. The transformation can be as simple as changing the space representation from $[x_1, x_2]$to $[x_1, x_2, x_1^2 + x_2^2]$. Explicitly mapping the vectors to the higher dimension might be computationally expensive. A kernel "trick" is employed to solve this problem.

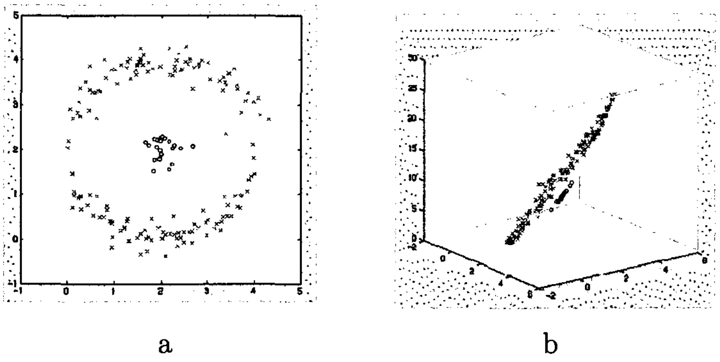


Figure 3.2: Classification at a higher dimension. a. 2 dimensions, b. 3 dimensions

3.1.1 Kernel trick

An in-depth analysis of kernel methods is given in [31, 32]. The first step is the choice of the kernel function, e.g.:

$$k(x, y) = (x.y)^d \tag{3.1}$$

This corresponds to a dot product in the space of d^{th} order monomials. Note that a value of $d = 1$ implements the linear form of PCA. Now we compute the kernel matrix,

$$K_{ij} = \left(k\left(x_i, x_j\right)\right)_{ij} = \left(\left(x_i . x_j\right)\right)_{ij} \tag{3.2}$$

where, x_i and x_j represent input patterns. For further analysis we want a normalized matrix. Normalizing in the higher feature space is not straight forward as in the input image space, but once we have the kernel matrix we can use the following equation to perform the normalization:

$$\tilde{K}_{ij} = K - I_M K - K I_M + I_M K I_M \tag{3.3}$$

where $I_M = I/M$, M is the number of images under consideration. The eigenvectors $\left(\alpha^1 \rightarrow \alpha^d\right)$ corresponding to the largest eigenvalues $\left(\lambda^1 \rightarrow \lambda^d\right)$ of the $\tilde{K}$ matrix give us the required principal components. The eigenvectors are normalized using the following condition,

$$1 = \lambda^k \left(\alpha^k . \alpha^k\right) \tag{3.4}$$

Principal components of a test point corresponding to the selected kernel are extracted using the following,

$$W_n\left(x_{test}\right) = \sum_{i=1}^{d} \alpha_i^n k\left(x_i, x_{test}\right) \tag{3.5}$$

3.1.2 Kernel functions

Selecting an optimal kernel and setting its various parameters is a problem in itself. Three different kernels were tried in [27], including the one in Equation 3.1. The other two were a Gaussian kernel and a sigmoid kernel as in the following equations,

$$k\left(x,y\right)=exp\left(-\left\|x-y\right\|^{2}/\sigma^{2}\right) \tag{3.6}$$

$$k\left(x,y\right)=\tanh\left(\left(x.y\right)+b\right)\ \text{ where } b\longrightarrow \text{bias} \tag{3.7}$$

The limiting requirement here is that a kernel should satisfy Mercer's theorem [33], resulting in positive semi-definite matrices. A polynomial function based kernel face recognition architecture on PCA is provided in [34]. Comparison of kernel approaches on eigenfaces and Fischer faces on kernels is given in [35]. An intra-personal probabilistic subspace [27] based approach is modified in [36] to a kernel based space to perform face recognition using a Gaussian kernel to project data to higher dimensional space. A discriminant approach on kernel space is provided in [37]. The relationship between kernel-PCA and MDS is explored in [38], with MDS being represented as a special case of kernel function. A kernel MDS approach is presented in [39], where MDS is applied to features extracted from FLD to form a kernel matrix. Kernels have also been formed from the neighborhood methods discussed below [40, 41].

3.2 Neighborhood dependent methods

Neighborhood dependent methods represent each input as some combination of its neighbors. The Isomap algorithm preserves the geometry of the data by finding the geodesic distance between all pairs of data points. On a nonlinear dataset, the shortest straight path between two points ignores the actual geometry of the dataset under consideration. Figure 3.3 illustrates this concept [10]. The dashed line represents the shortest straight distance and the solid line represents the geodesic distance.

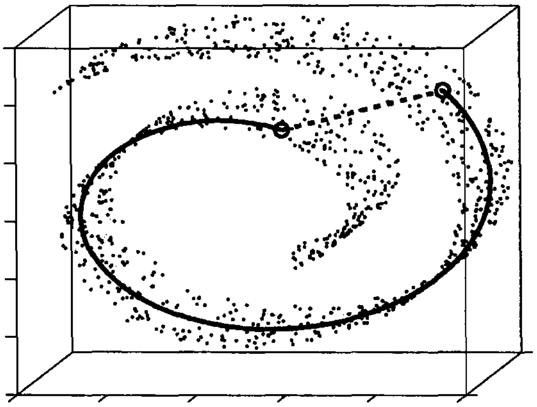


Figure 3.3: Illustration of geodesic distance.

3.2.1 Isomap

The first step in the Isomap process is to find the Euclidean distances between all data points and to represent these values in the form of a square matrix. Each value in the matrix represents the distance between the pattern indexed in the row to the corresponding column indexed pattern. Once the Euclidean distance matrix is formed out of the input data points, finding the geodesic distance [42] involves computing a minimal cost trajectory from each point to the other points. For a given data point, length of the shortest path to another point is obtained by summing the length of the segments along the shortest path. This value roughly equates to true geodesic distance between two points. Collecting all the details together, Isomap goes through the following steps (Floyd's [43] algorithm is used to compute geodesic distances here),

1. Euclidean distance measure, D_x, between all input points is obtained.

2. Based on the distances $d_x(i,j)$ between points i and j a neighborhood is defined for each point. This could be done by defining a distance threshold ϵ or, in case of k nearest neighbor method, checking if i is one of the k nearest neighbors of j. All the edge lengths are set to $d_x(i,j)$.

3. Initialize the geodesic distance, $d_G(i,j) = d_x(i,j)$ if i,j are linked by an edge, else $d_G(i,j) = \infty$.

4. Update $d_G(i,j) = min\left\{d_G(i,j), d_G(i,j) + d_G(k,j)\right\}, 1 \leq k \leq M$.

5. Center the distance matrix D_G, $D'_G = \frac{1}{2}HD_G^2H$.

   where, $H = I - \frac{1}{N}\left[1_1 \ldots 1_N\right]^T \left[1_1 \ldots 1_N\right]$

6. Compute eigenvalues (λ) and eigenvectors (v) of the centered matrix.

7. Define projected point as: $y^i = \sqrt{\lambda}v^i$

M is the number of points in the dataset. The method is used for face recognition in [13]. Selection of neighborhood parameters given in step 2 of the Isomap algorithm cannot be made arbitrarily as this affects the performance of the algorithm.

If the neighborhood is large, points from other branches of the manifold could be erroneously included, leading to the problem of short-cuts [44]. The most successful implementation of Isomap for face recognition is found in [45, 46], where a linear discriminants based analysis is used instead of PCA in the projection step.

3.2.2 Extended-Isomap

Isomap as applied to multiple classes may not produce the most discriminating features. Extended-Isomap utilizes Fischer Linear Discriminants (FLD) method to the Isomap method for pattern classification. The first step in this method is to follow the original Isomap method and find the geodesic distance of each point to every other point. Every point is now represented by a feature vector that is based on this distance measure. FLD is now applied to the feature vectors to find an optimal projection direction for classification. Between-class and within-class scatter matrices are found by,

$$S_B = \sum_{i=1}^{C} N_i \left(\mu_i - \mu\right)\left(\mu_i - \mu\right)^T \tag{3.8}$$

$$S_W = \sum_{i=1}^{C} \sum_{f_k \in Z_i} \left(f_k - \mu_i\right)\left(f_k - \mu_i\right)^T \tag{3.9}$$

where, f_k are the data samples, μ is the mean of all samples of f_k, μ_i is the mean of class Z_i with N_i samples and C is the total number of classes. The optimal projection is found by solving the function,

$$W_{FLD} = arg\left\{\underset{W}{max}\frac{|W^T S_B W|}{|W^T S_W W|}\right\} \tag{3.10}$$

by finding the generalized eigenvectors of S_B and S_W, corresponding to the largest generalized eigenvalues.

One of the problems that occurs here is the possible singularity of the within-class scatter matrix. One way to overcome this is to add a multiple (ε) of the identity matrix to this matrix ($S_W + \varepsilon I$). Another approach is to use PCA and work within a reduced dimension [47]. Combining the two methods can also result in improved face recognition accuracy. A theoretical base for combining PCA and LDA is given in [48]. A representation of face images combining FLD with MDS can be seen in [49], where MDS is applied to features extracted using FLD. MDS also forms the base for some of the nonlinear feature extraction methods discussed later.

3.2.3 Other nonlinear approaches

An approach very similar to using geodesic distance uses curvilinear distance analysis presented in [50]. Another method that tries to preserve the geometry of the original dataset is the Locally Linear Embedding (LLE) method given in [11], where a linear slope line is estimated using least squares in a specified neighborhood. Multiple straight lines are modeled to cover the entire dataset. A reconstruction error cost function defined by,

$$E(W) = \sum_i \left| x_i - \sum_j W_{ij} x_j \right|^2 \tag{3.11}$$

where, x is a pattern vector, i, j represent the data points and W_{ij} is the coefficient relating points i and j. A further illustration of nonlinearity of face images can be seen from Figure 3.4, where the points spread out based on pose or expression variation [11].

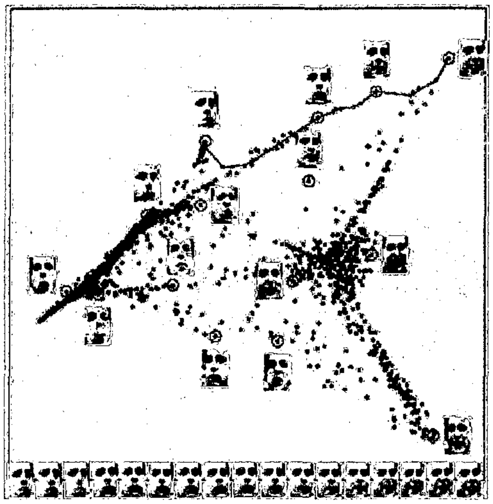


Figure 3.4: Faces on first two LLE coordinates.

Laplacian eigenmaps method [51] defines relationship between neighboring points by edges. The edges are defined by either a 1 (connected) or a 0 (disconnected), or in a more complex representation by an exponential. Eigenfaces method [52] uses this concept of neighborhoods to model face data and applies it to face recognition. A comparison of Isomap, LLE and Laplacian eigenmaps methods is given in [53]. That paper also presents a modified Isomap method where edge weights are adjusted by a mean distance factor. A MDS based approach to nonlinear manifolds is presented in [54] where MDS is applied to obtain a reduced feature space followed be a neighborhood dependent piece-wise linear approach to obtain the manifold. The principal curves method [55] models data as a smooth polynomial curve. The method is used for face recognition in [56] illustrating the advantage of using nonlinear curves instead of linear slopes.

3.3 Summary

The issue with the nonlinear methods presented here is that the information content is in neighborhood defined format, disabling a way to retrieve original data components. Another general issue with these methods is their batch mode of operation, making them disadvantageous when it comes to fast face recognition. A new test point will have to go through all the steps to be projected on to the new subspace. Also, the introduction of the new point changes the neighborhood for points already in the database. For a fast face recognition system, a direct projection from the image space to the nonlinear subspace is needed. We move forward with the general idea of combining the nonlinear concepts of these methods to the direct projection ability of PCA to obtain a nonlinear subspace projection methodology. We propose here a method where these nonlinearities are modeled using a least squares estimation approach and the images are now projected on to a low dimensional, multi-ordered nonlinear region. This ensures that we retain the geometric properties of the dataset in the low dimensional feature space. A new pattern is classified by projecting to the low dimensional nonlinear region and computing shortest distances to the trained patterns inside the regions. The approach of using discriminants is implemented by using nonlinear subspace projection method to reduce the original dimensionality of the data and then performing discriminant projection on the resulting features.

4 Subspace Projection Method for Nonlinear Manifolds

A method for obtaining the shape and structure of a nonlinear manifold in the feature space representing a family of patterns is presented in this section. The parameters defining the manifold are obtained by analyzing the characteristics of the original data. A nonlinear transformation methodology is developed to project a pattern in the original space on to the respective position in the feature space. A face recognition system is developed based on this approach for extracting facial feature components. Face recognition accuracy is improved by following a discriminant approach to obtain well distinguishable clusters in the nonlinear subspace. A modular technique is applied on the images to better model local data using nonlinear subspaces.

4.1 Projection to a curved manifold

The basis for the proposed method lies in the work done by Seow et al. [14], where memory association is represented using lines of attraction. The memory is modeled using a recurrent neural network and stored memory is obtained by convergence. We extend this idea onto classical component analysis.

Consider a neural network with two layers of neurons as shown in Figure 4.2. Each neuron is considered as a nonlinear combiner. The relationship of each neuron with respect to every other neuron is expressed as a k-order polynomial for stimulus-response pair a^s, b^s corresponding to the s^{th} pattern given by:

$$b_i^s = \sum_{j=1}^{L} \sum_{m=0}^{k} w_{(m,ij)}^s (a_j^s)^m \, \forall i \tag{4.1}$$

where, $L = N^2$ for $N \times N$ image, $w_{m,ij}^s$ is the m^{th} order weight value between neurons i and j as shown in Figure 4.1.

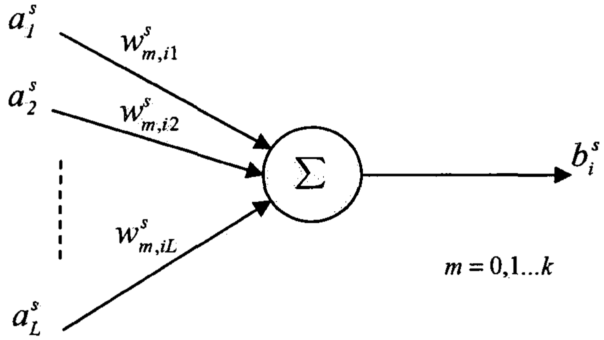


Figure 4.1: Output neuron model.

The resulting m^{th} order weight matrix is expressed as,

$$W_m^s = \begin{bmatrix} w_{m,11}^s & \cdots & w_{m,1L}^s \\ \vdots & \ddots & \vdots \\ w_{m,L1}^s & \cdots & w_{m,LL}^s \end{bmatrix} \tag{4.2}$$

The above discussion leads to modification of the covariance between the variables by representing them as coefficients of a k^{th} order curve instead of a linear slope value, resulting in multiple covariance matrices C_m.

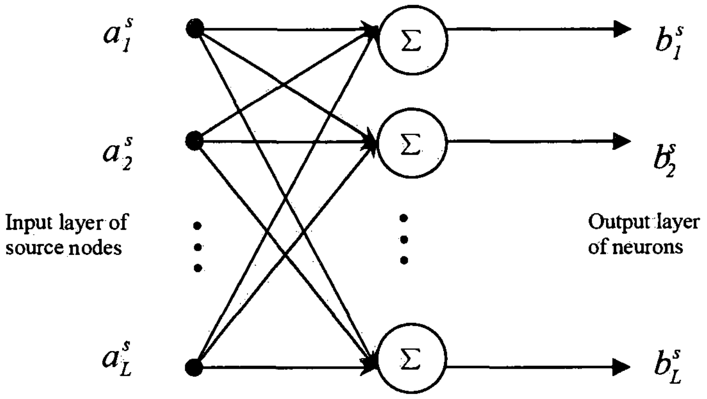


Figure 4.2: Memory model.

$$C_m^s = \begin{bmatrix} c_{m,11}^s & \cdots & c_{m,1L}^s \\ \vdots & \ddots & \vdots \\ c_{m,L1}^s & \cdots & c_{m,LL}^s \end{bmatrix} \text{ for } 0 \leq m \leq k \tag{4.3}$$

Covariance is a measure of how much two variables change together. Covariance is positive when two variables change together. Now in this particular case, each image is an observation and each pixel in an image is a variable under consideration. The covariance matrix is a collection of covariances between all the variables under consideration. So define an image vector I in a set of M images as,

$$I = \begin{bmatrix} i_1 \\ \vdots \\ i_{N^2} \end{bmatrix} \tag{4.4}$$

where i_i is an image pixel. The mean image vector of all the M images in the set can be found by,

$$Y = \frac{1}{M}\sum_{j=1}^{M} I_i \rightarrow Y = \begin{bmatrix} y_1 \\ \vdots \\ y_{N^2} \end{bmatrix} \tag{4.5}$$

The centered image A can be obtained as,

$$A = I - Y = \begin{bmatrix} a_1 \\ \vdots \\ a_{N^2} \end{bmatrix} \tag{4.6}$$

Define a product matrix P as,

$$P^s = A^s A^{sT} \tag{4.7}$$

where A is the centered image vector of size $L = N^2$. Combining the product matrices of Equation 4.7 to form k covariance matrices boils down to a curve fitting problem. The least squares estimation method finds the values of the constants in the chosen equation that minimize the sum of the squared deviations of the observed values from those predicted by the equation. This can be represented in equation form as,

$$J(c) = J(c_{0,ij}, c_{1,ij} \ldots, c_{k,ij}) = \sum_{s=1}^{M} \left[p_{ij}^s - \sum_{m=0}^{k} c_{m,ij} \left(a_j^s\right)^m \right]^2 \tag{4.8}$$

where M is the number of patterns and p_{ij} are elements of the product matrix P. The elements of product matrices given by Equation 4.7 form the observed values. The necessary conditions for Equation 4.8 to minimize are that the partial derivatives

$\frac{\partial J}{\partial c_m} = 0$ for $m = 0, 1, \ldots, k$. This gives us $k+1$ equations to find the $k+1$ coefficients of the k^{th} order polynomial,

$$\frac{\partial J\left(c\right)}{\partial c_m} = 0 = 2\sum_{s=1}^{M} \left(a_j^s\right)^m \left[p_{ij}^s - \sum_{m=0}^{k} c_{m,ij} \left(a_j^s\right)^m\right] \tag{4.9}$$

The above set of Equations 4.9 result in a unique solution which gives a minimum for $J(c)$. Coefficients $c_{m,ij}$ approximate a curve on M pattern points. This method is illustrated in Figure 4.3 where product matrix coefficients relating positions $1, L$ of different samples from the same class are combined using least squares estimation to obtain the covariance values at position $1, L$.

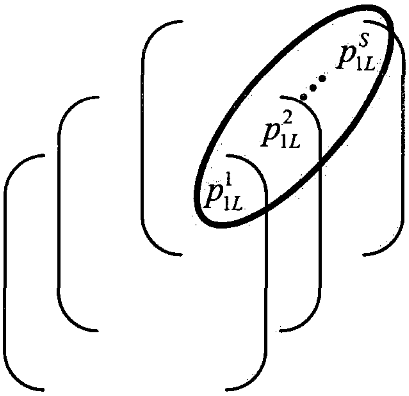


Figure 4.3: Fitting curves to product matrix data.

The discussion so far has dealt with face images still being considered as a high dimensional vector. Our main aim, though, is to project this image onto a lower dimensional nonlinear manifold. A set of matrices E_m consisting of eigenvectors that diagonalize the set of covariance matrices C_m is computed; thus,

$$E_m^{-1} C_m E_m = D_m \; for \; 0 \leq m \leq k \tag{4.10}$$

where D_m is the diagonal eigenvalue matrix of C_m. The subset of eigenvectors corresponding to the largest $d\,(d \ll N^2)$ number of eigenvalues is selected as the basis of the data. The d dimensional data can be expressed as,

$$\begin{bmatrix} b'_1 \\ \vdots \\ b'_d \end{bmatrix} = \begin{bmatrix} E_{k,11} & \cdots & E_{k,1L} \\ \vdots & \ddots & \vdots \\ E_{k,d1} & \cdots & E_{k,dL} \end{bmatrix} \begin{bmatrix} (a_1)^k \\ \vdots \\ (a_L)^k \end{bmatrix} + \cdots + \begin{bmatrix} E_{0,11} & \cdots & E_{0,1L} \\ \vdots & \ddots & \vdots \\ E_{0,d1} & \cdots & E_{0,dL} \end{bmatrix} \begin{bmatrix} (a_1)^0 \\ \vdots \\ (a_L)^0 \end{bmatrix} \tag{4.11}$$

The effect of the projection is illustrated in Figure 4.4 in a two dimensional sense. Compared to a line as in classical linear methods, projection onto a curve is expected to give a more accurate representation of the position of an image in the low dimensional space.

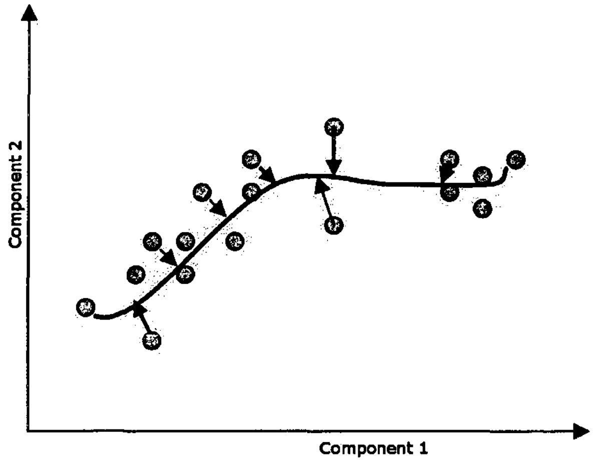


Figure 4.4: Projection to curve.

The linear PCA method is considered here as a starting point. PCA aims to encode relevant information in a face image and represent the image as a low dimensional vector. The variation information in a collection of face images is used to encode and compare individual face images. The principal components or the eigenvectors of the covariance matrix of the set of face images are found. Each image in the data set is now represented using a linear combination of the principal components. Here we can approximate the set by using only the best possible eigenvectors corresponding to the first few of the largest eigenvalues. Consider the face images to be of size N by N. These images are represented by a vector of size N^2. Since the face images have similar structures, the vectors representing them will be correlated. These vectors define the subspace of the face images which is called the "face space". Due to the correlation, the images can be represented in a lower dimensional space. The covariance matrix of a set of M images of size $N \times N$ is formed by finding an average product matrix as,

$$C = \frac{1}{M} \sum_{s=1}^{M} P^s \tag{4.12}$$

The point representation or the linearity of the method is a result of this averaging. By using a mean value strategy to combine the various patterns, we are forcing one value to represent the connection between pixels which is representative of the slope of a line.

The variance between the first two pixels can be calculated as,

$$c_{12} = E\left((i_1 - y_1)\left(i_2 - y_2\right)\right) \tag{4.13}$$

$$c_{12} = \frac{1}{M} \sum_{j=1}^{M} a_1^j a_2^j \tag{4.14}$$

We can see where the average product matrix in Equation 4.12 comes from. Equation 4.14 can be looked at from another point of view, as the slope for a straight line. The best way to put this would be to describe c_{12} as the slope of the line through which the two variables vary. Thus, the covariance matrix C is a collection of slopes that defines how variables change together. This idea can be made clearer using the following discussion [1].

The main idea behind component analysis is to find a single vector representation for all the M images. The simplest representation is the mean of all images, but this just gives another point in the image space and fails to show any variability in the image data. Variability can be obtained by projecting the data onto a line running through the sample mean. Let E be a unit vector in the direction of the line. The equation of the line can now be written as,

$$I = Y + \alpha E \tag{4.15}$$

where the scalar α is a real value and corresponds to the distance between the mean Y and any image I. Equation 4.15 can be generalized as,

$$I_j = Y + \alpha_j E \tag{4.16}$$

An optimal set of values for the coefficients α_j can be found by minimizing the squared error criterion,

$$J(\alpha,\, E) = \sum_{k=1}^{M} \left\| (Y + \alpha_j E) - I_j \right\|^2 \tag{4.17}$$

This gives the solution for α_j as,

$$\alpha_j = E^T (I_j - Y) \tag{4.18}$$

How does all this relate to the covariance matrix though? This can be seen by expanding Equation 4.17 as,

$$\begin{aligned} J(\alpha,\, E) &= \sum_{j=1}^{M} \alpha_j^2 \left\| E \right\|^2 - 2 \sum_{j=1}^{M} \alpha_j E^T (I_j - Y) + \sum_{j=1}^{M} \left\| I_j - Y \right\|^2 \\ &= \sum_{j=1}^{M} \alpha_j^2 - 2 \sum_{j=1}^{M} \alpha_j^2 + \sum_{j=1}^{M} \left\| I_j - Y \right\|^2 \\ &= -\sum_{j=1}^{M} \left[E^T (I_j - Y) \right]^2 + \sum_{j=1}^{M} \left\| I_j - Y \right\|^2 \\ &= -\sum_{j=1}^{M} E^T (I_j - Y)(I_j - Y)^T E + \sum_{j=1}^{M} \left\| I_j - Y \right\|^2 \\ &= -ME^T CE + \sum_{j=1}^{M} \left\| I_j - Y \right\|^2 \end{aligned} \tag{4.19}$$

Any vector E that minimizes Equation 4.17 also maximizes $E^T CE$ in Equation 4.19. This criterion sets E to be the eigenvectors corresponding to the largest eigenvalue of the covariance matrix. Thus, we are projecting the images to a line, passing through the mean, corresponding to the largest eigenvalue. This is only a one dimensional representation. The idea can be extended to d dimensions by selecting d eigenvectors corresponding to the largest d eigenvalues, where $d \ll N^2$.

A linear discriminant function on vector A can be written as,

$$B = \omega_0 + W^T A \tag{4.20}$$

where W is the weight vector and ω_0 the bias. The equation $B = 0$ defines a decision surface that separates points assigned to two classes. The surface formed here is a hyperplane. Figure 4.5 shows a linear classifier having N input units, each corresponding to the values of an input vector. Each input value is multiplied with its corresponding weight w_{ij}, $1 \leq i, j \leq N$ (j^{th} input to the i^{th} neuron). The variable s is an index to represent input training patterns. Multiple classes can be handled by having multiple linear discriminants leading to a more generalized equation,

$$b_i = \omega_{i0} + W_i^T A \text{ where } 1 \leq i \leq N \tag{4.21}$$

The above equation divides the region into N regions.

The discriminant function given in Equation 4.21 can be expanded as,

$$b_i = \omega_{i0} + \sum_{j=1}^{N} w_{ij} a_j \tag{4.22}$$

where the coefficients w_{ij} are the components of weight vector W. A quadratic discriminant function can be obtained by adding additional terms to Equation 4.22:

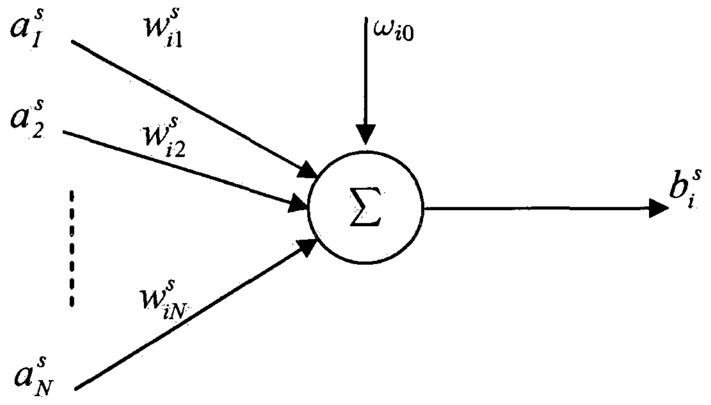


Figure 4.5: Linear discriminant.

$$b_i = \omega_{i0} + \sum_{j=1}^{N} w_{ij}a_{ij} + \sum_{j=1}^{N}\sum_{l=1}^{N} w_{i\bar{j}l}a_j a_l \tag{4.23}$$

where $w_{i\bar{j}l}$ is the weight value from j^{th} and l^{th} input combination to i^{th} neuron. The quadratic function has additional coefficients to create complicated surfaces. Arbitrary terms from Equation 4.23 can be chosen to form polynomial discriminant functions. The quadratic equation can be expanded to higher dimensions.

The weight coefficients can be calculated using learning rules. The simplest form of weight selection is the Hebbian learning described as [1],

$$w_{ij} = \frac{1}{N}\sum_{s=1}^{P} a_i^s a_j^s \tag{4.24}$$

Equation 4.24 is very similar to Equation 4.14 for covariance. The weight coefficients and the covariance coefficients are related by a scalar, showing the relation between the two methods.

4.2 Modular approach

The modular approach works by dividing the images into a number of sub-images. Projection weight vectors are computed for each of the sub-images and concatenated to form the representative weight vector for the original image. When there is a variation in illumination or expression, only some of the sub-images vary, and the rest of the face regions will remain the same as the original image [29]. This was modified to consider each of the sub-images separately and have separate covariances and thus, projections to separate image spaces in [30, 57]. Both approaches were found to result in improved accuracy over global approaches. Each image in the training set is divided into R sub-images of size $^{N}/_{\sqrt{R}} \times {}^{N}/_{\sqrt{R}}$. The j^{th} sub-image of the image I_i can be obtained as,

$$I_{ij}\left(m,n\right) = I_i\left(\frac{N}{\sqrt{R}}\left(j-1\right)+m,\ \frac{N}{\sqrt{R}}\left(j-1\right)+n\right)\ for\ 1 \leq m,n \leq \frac{N}{\sqrt{R}} \tag{4.25}$$

where, $1 \leq i \leq M$ and $1 \leq j \leq R$. The average image (Y_j) for each sub-image is computed and each sub-image is normalized (A_{ij}) as,

$$Y_j = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{R} I_{ij} \tag{4.26}$$

$$A_{ij} = I_{ij} - Y_j \tag{4.27}$$

Separate covariance matrices are obtained for each sub-image set as,

$$C^s_{mj} = \begin{bmatrix} c^s_{mj,11} & \cdots & c^s_{mj,1(L/R)} \\ \vdots & \ddots & \vdots \\ c^s_{mj,(L/R)1} & \cdots & c^s_{mj,(L/R)(L/R)} \end{bmatrix} \tag{4.28}$$

The idea can be grasped from Figure 4.7, where the network is modified for modularity with 4 sub-images. The number of nodes is now reduced by a factor of 4 by increasing the number of networks to 4. Eigenvectors corresponding to d largest eigenvalues are computed for each of the covariance matrices.

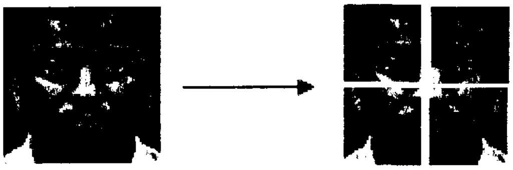


Figure 4.6: Sub-image formation.

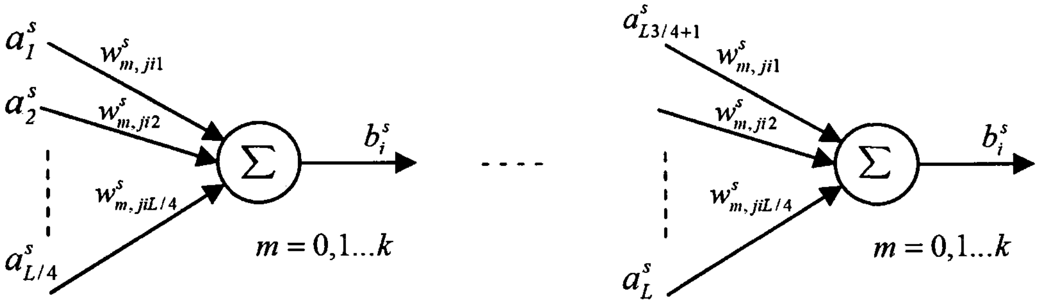


Figure 4.7: Modular network.

Weight vectors are computed for each sub-image as,

$$B_{ij} = \sum_{m=0}^{k} E^T_{mj} \left(I_{ij} - Y_j\right)^m \tag{4.29}$$

where, E_{mj} are the eigenvectors corresponding to the d largest eigenvalues of C_{mj} and

m represents the order of the polynomial. A test image can be transformed to the feature space as,

$$B_{testj} = \sum_{m=0}^{k} E_{mj}^{T} \left(I_{testj} - Y_j \right)^m \tag{4.30}$$

4.3 Discriminant analysis

Components found by PCA might not be the most useful in discriminating data in different classes [1]. Samples from well formed clusters, when projected onto an arbitrary line, can produce a confused mixture of samples from all the classes leading to poor recognition performance. Discriminant analysis aims to find an optimal line orientation for which the data classes are well separated. Discriminant analysis is widely applied for face recognition and has been found to perform better or as well as PCA [58]. The discriminant method has been applied to the manifold model in the extended Isomap method [45] and resulted in improved accuracy over a straight Isomap implementation that was based on PCA. Fischer Linear Discriminants (FLD) considers maximizing the following objective,

$$J\left(W\right) = \frac{W^T S_B W}{W^T S_W W} \tag{4.31}$$

where S_B is the between-class scatter matrix and S_W is the within class scatter matrix. These scatter matrices can be defined as,

$$\begin{aligned} S_B &= \sum_{l=1}^{C} (\mu_l - \mu)(\mu_l - \mu)^T \\ S_W &= \sum_{l=1}^{C} \sum_{I \in C_l} (I - \mu_l)(I - \mu_l)^T \end{aligned} \tag{4.32}$$

where I is an image vector of size N^2, μ is the mean of all data, μ_l is the mean of the l^{th} class C_l and C is the total number of classes. The within-class scatter matrix gives the covariance of data in each class. This matrix can now be modified for the l^{th} class by using the k^{th}order polynomial representation to get,

$$S_{Wm}^{l} = \begin{bmatrix} s_{m,11}^{l} & \cdots & s_{m,1L}^{l} \\ \vdots & \ddots & \vdots \\ s_{m,L1}^{l} & \cdots & s_{m,LL}^{l} \end{bmatrix} \quad \text{where} \quad 0 \leq m \leq k,\; L = N^2,\; 1 \leq l \leq C \tag{4.33}$$

The least squares equation now looks like,

$$J\left(s_{0,ij}^{l}, s_{1,ij}^{l}, \ldots, s_{k,ij}^{l}\right) = \sum_{s=1}^{M_l} \left[P_{ij}^{s} - \sum_{m=0}^{k} s_{m,ij} \left(a_j^s\right)^m \right]^2 \tag{4.34}$$

$$s_{0,ij}, s_{1,ij}, \ldots, s_{k,ij} = \sum_{l=1}^{C} s_{0,ij}^{l}, s_{1,ij}^{l} \ldots, s_{k,ij}^{l} \quad for \quad 1 \leq i, j \leq L \tag{4.35}$$

where M_l is the number of images in the l^{th} class. The product matrix is given by,

$$P^l = \left(I^l - \mu_l\right)\left(I^l - \mu_l\right)^T \tag{4.36}$$

During the implementation of the method, there were issues with the dimensionality of the data used and the number of samples available. To solve this, a method similar to the one followed in [59] was used, where the image was first projected on to the face subspace using PCA and FLD was used to obtain a linear classifier. Instead of linear methods, nonlinear manifold based projections were used in both cases.

4.4 Phase congruency

Image features like edges result in points where the frequency components are maximally in phase. Phase congruency features have been used to good effect for face recognition in [60] by following a modular approach on feature images. The technique followed by Kovesi in [61] is used here. The phase congruency function is defined as,

$$PC(x) = \max_{\theta \in [0, 2\pi]} \frac{\sum_n A_n cos\left(n\omega x + \phi_n - \theta\right)}{\sum_n A_n} \tag{4.37}$$

where A_n is the amplitude of the n^{th} Fourier component, ω is a constant (usually 2π) and the value of θ that maximizes Equation 4.37 is the weighted mean of all the Fourier terms at the point being considered. To actually compute the features, peaks in a local energy function are considered.

$$E\left(x\right) = \sqrt{F^2\left(x\right) + H^2\left(x\right)} \tag{4.38}$$

where $H(x)$ is the Hilbert transform of $F(x)$. This function is related to $PC(x)$ as,

$$E(x) = PC(x)\,\Sigma_n A_n \tag{4.39}$$

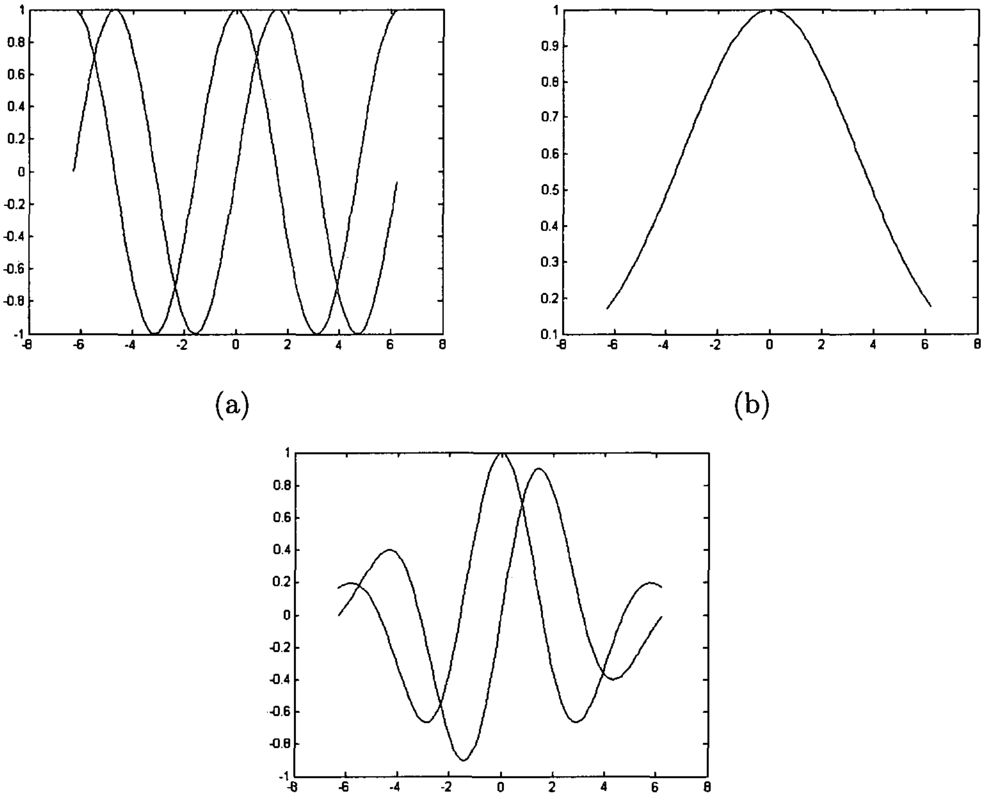
Thus, the local energy function is directly proportional to the phase congruency function, so, local energy peaks should relate to peaks in phase congruency. Instead of a direct computation, the local energy of an input image can be calculated by convolving with a pair of filters in quadrature. The signal F is obtained by removing the DC component of input image I. The new image is saved and convolved with another filter that is in quadrature with the first filter. The result of the two convolutions are squared and saved. Another way to do this is to use log Gabor functions. Two wavelets are defined, one odd (M_n^o) and one even (M_n^e). These are obtained by modulating sine and cosine waves by a Gaussian. All three signals involved are shown in Figure 4.8, the sine and cosine signals in Figure 4.8a are modulated by the Gaussian in Figure 4.8b to get the odd and even wavelets in Figure 4.8c.

The amplitude and phase component of the Fourier component at a given scale n are,

$$A_n(x) = \sqrt{(I(x) * M_n^e)^2 + (I(x) * M_n^o)^2} \tag{4.40}$$

$$\phi_n(x) = \tan^{-1}(I(x) * M_n^e / I(x) * M_n^o) \tag{4.41}$$

where, " $*$ " represents convolution. The signals F and H can be obtained by,



(a)

(b)

(c)

Figure 4.8: (a)Sine and cosine waves, (b)Modulating Gaussian, (c)Symmetric wavelets.

$$\begin{aligned} F(x) &= \Sigma_n I(x) * M_n^e && (4.42) \\ H(x) &= \Sigma_n I(x) * M_n^o && (4.43) \end{aligned}$$

Thus, the sum of amplitudes of the frequency components is given by,

$$\sum_n A_n(x) = \sum_n \sqrt{(I(x) * M_n^e)^2 + (I(x) * M_n^o)^2} \qquad (4.44)$$

The expression for phase congruency is slightly modified adding a small positive constant to the denominator to prevent the Equation from becoming unstable when $\sum_n A_n(x)$ and $E(x)$ becomes small. Thus we have,

$$PC(x) = \frac{E(x)}{\varepsilon + \sum_n A_n(x)} \tag{4.45}$$

One dimensional analysis is carried out over several orientations, and the results are combined to obtain a two dimensional image. Some images together with their phase congruent output are shown in Figure 4.9.

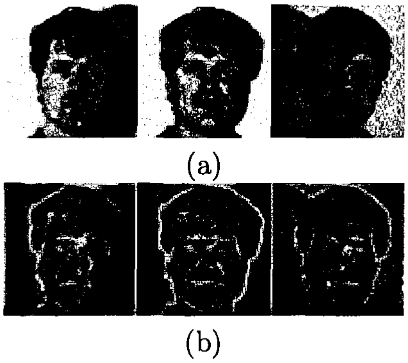


Figure 4.9: (a) Original images, (b)Phase congruent images.

Phase congruent features are dimensionless quantities and are invariant to changes in image brightness and contrast [62]. This enables phase congruent features to better represent the features in an image. This better representation of an image should help in an application like face recognition where we are trying to classify faces based on the detected features. The approach should particularly help in overcoming problems with illumination variations. The obtained edge images are now used to train the nonlinear subspace components. A new input image is transformed to the phase congruent edge image and projected on to the feature space before classification.

4.5 Reconstruction from nonlinear principal components

Recovering the original image from the principal components is important to emphasize the authenticity of the image representation method presented in this dissertation. An attempt is made to reconstruct the original image from the average image and the principal components obtained by the proposed dimensionality reduction technique. For a k^{th} polynomial based nonlinear function, we can reconstruct the image as,

$$I = Y + \sum_{m=0}^{k} E_m^T B_m \tag{4.46}$$

where, E is the set of eigenvectors corresponding to a few of the largest eigenvalues, B is the reduced dimension weight vector for the image under consideration and m denotes the order of the component.

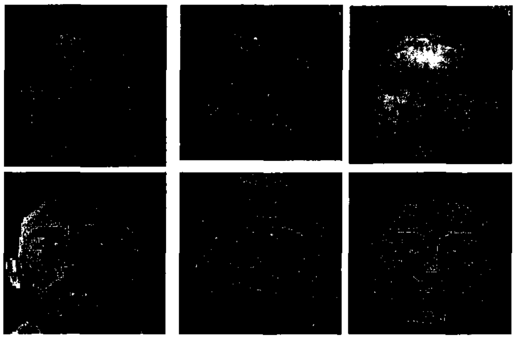


Figure 4.10: Face reconstruction from nonlinear space. Top row: Reconstructed faces, Second row: Original faces.

The first row in Figure 4.10 shows the result of face reconstruction from the nonlinear face subspace of the original images from the ORL face database [63] (shown in Figure 4.10 second row). Image features in the nonlinear subspace lie inside a region defined by the nonlinear function selected to model the space. A polynomial function is used in this dissertation. To reconstruct the data, we use the same nonlinear function. Feature components from each separate ordered space are now multiplied with the principal components to obtain the high dimensional representation for that particular subspace. This representation gives only centered values which are now integrated with the mean image (Y) obtained during training to obtain the final reconstructed face.

5 Experimental Results and Discussion

Four different face databases covering four different criteria are considered for testing the proposed subspace based methodology for feature extraction. The Yale face database [64] contains images with varying expressions and lighting conditions. Performance of the algorithms on a large set of faces can be tested using the FERET database [65, 66]. FERET also provides a standard testing protocol and previously reported set of images for training and testing purposes. The effects of varying illumination are noted using a subset from the illumination set of the PIE face database [67]. The effect of varying expressions are illustrated using the CMU expression variant face database [68]. The testing environment used for evaluating the proposed face recognition technique is illustrated in Figure 5.1. The various steps in the process can be listed as,

1. Face images are divided into training and testing sets. The specific way in which this was done for each database is given in the respective subsections. The selected trained images undergo histogram equalization as a pre-processing step.

2. Images are divided into four subimages in the modular approach, and the manifold projection technique is applied to each subimage.

3. Each image is represented in vector form by appending subsequent row pixels.

4. A mean image vector of all the training image vectors is obtained. The

training images are now centered by subtracting the mean from each image. For the discriminant approach, class means are also calculated by finding means of images in the same class.

5. Covariance matrices corresponding to a second order polynomial function are obtained using Least Squares Estimation (LSE).

6. Principal components are obtained by eigen analysis of the covariance matrices. Eigenvectors corresponding to the first few largest eigenvalues are selected. The number of eigenvectors selected determines the dimension of the feature space.

7. Representative feature vectors are now obtained for each training image by transformation using the second order polynomial equation. Between-class and within-class scatter matrices are determined in this reduced dimension space. Discriminant features are obtained by eigen analysis of these two matrices and projecting the data using these components.

8. Test images are centered using the mean image. Four subimages are formed for the modular approach. A test image feature vector is formed by using the trained principal components or discriminant components.

9. Euclidean distances are found between feature vectors of the test image and those of the training images. A nearest neighbor strategy is used for determining the identity of the test face.

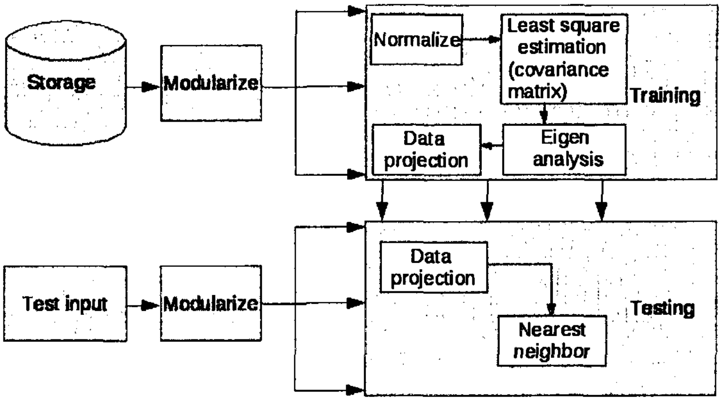


Figure 5.1: Test system block diagram.

5.1 Yale face database

The proposed algorithm was run on the Yale face database containing 165 images of 15 individuals. The images are taken under a set of varying illumination conditions and expressions. There are images with normal, sad, happy, surprised and winking expressions. There are also images where the position of the light source is at the center, right or left. The presence of eye glasses in some images also gives variability in occlusion. Figure 5.2 gives some sample images from the dataset. Each image was of size 64 × 64 pixels, and no modifications were made to the images. A leave one out testing strategy was followed. One image of each person is left out of the training sample. The images that were left out are then used as test images. This enables us to test the effect of each variation on the accuracy of the method.

Figure 5.3 shows the results obtained for the proposed method versus results for baseline PCA for this database. The number of dimensions of the feature space was varied and recognition accuracy for each dimension was noted. The two values were then plotted against each other to obtain the curves. The proposed method showed improved accuracy in all dimensions while large improvement in accuracy was obtained in lower dimensions. This illustrates the better ability of the approach to model variations in the input image data. From Figure 5.3, we can see that the recognition rate goes flat above 20 dimensions.

Figure 5.2: Yale face database sample images.

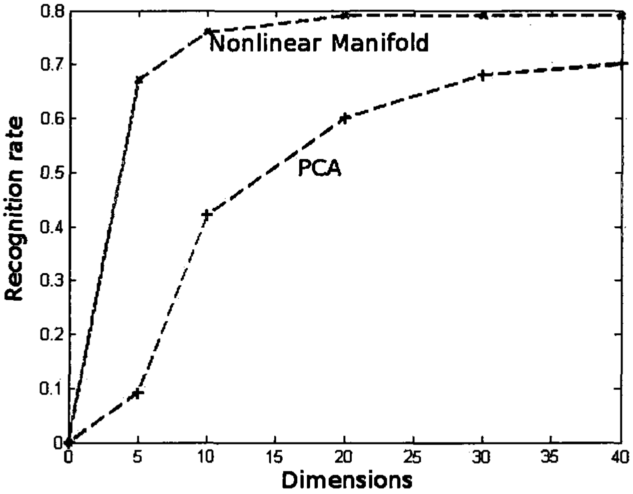


Figure 5.3: Accuracy curve for Yale database.

Table 5.1 compares recognition accuracy for the proposed nonlinear subspace method with results for other nonlinear manifold based methods (Isomap, LLE and Extended-Isomap presented in [45]). The leave-one-out testing strategy is followed in all the methods. PCA results are shown as a baseline to account for variations occurring due to the image sizes considered and normalizing steps involved. We can see that the results from the proposed method are better than straight nonlinear manifold approaches like Isomap and LLE. The accuracy for Isomap and LLE based methods is also dependent on the size of the neighborhood selected, as can be seen from the two results for the Extended-Isomap method. Further results presented in [45] for varying neighborhood sizes strengthen this view, the results provided here being the ones reported as the best. Among the proposed methods, the nonlinear discriminant approach gave the best results, illustrating the advantage of using class information.

Table 5.1: Accuracy values on Yale database.

Method	Accuracy(%/#)
PCA, 30 *dimensions*	68.4(113/165)
PCA [45], 30 *dimensions*	71.51(118/165)
Isomap, $k = 50$	71.51(118/165)
LLE, $k = 10$	73.93(122/165)
Extended-Isomap, $k = 25$	78.8(130/165)
Extended-Isomap, $\varepsilon = 12$	90.3(149/165)
Nonlinear subspace approach	**79.39(131/165)**
Modular approach	**82(135/165)**
Nonlinear discriminants	**84(138/165)**

5.2 CMU PIE face database

A sub-set of 11 images of 15 individuals for varying illumination from the CMU Pose, Illumination and Expression (PIE) [67] database is considered in this experiment. Sample images from the subset considered are shown in Figure 5.4.

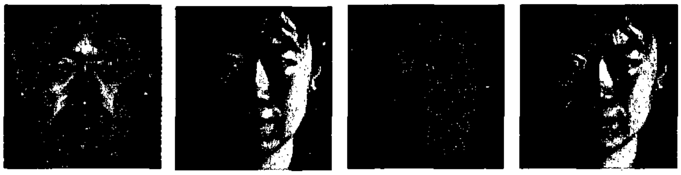


Figure 5.4: Sample images from the PIE illumination database.

The position of the light source varies from left to right in several steps. This database provides a much more detailed illumination variation as compared to the above Yale database. Also there is no variation in expression of the individual. This gives a better idea of accuracy of each method relative to illumination variation. The face images were cropped and resized to 64 × 64 pixels. The subset of images is divided into two groups based on brightness levels. All the dark images were grouped into one and the rest formed the other group. One set is used for training and the other is used for testing purposes. The procedure is repeated by interchanging the training and testing sets. The number of dimensions in the feature space is set to 20 and the result is shown in Table 5.2, illustrating the ability of the proposed method to model varying illumination better than the classical PCA method.

An increase in accuracy was found by following histogram equalization on both training and testing images as illustrated in Table 5.3. Enhancement of the set of images reduces variability of the face images due to illumination thus enabling simpler

Table 5.2: Accuracy values on PIE illumination database.

Method	Accuracy(%/#)
PCA	83.6(138/165)
Nonlinear Discriminants	94.54(156/165)

curves to model the data. Enhancing the test images enables reduced variability of the test faces from the trained samples, leading to better positioning in feature space near to their respective classes. This further leads to easier classification and better accuracy.

Table 5.3: Effect of histogram equalization on nonlinear discriminants.

Method	Accuracy(%/#)
Histogram equalized	98.1(162/165)
Raw images	94.54(156/165)

5.2.1 Effect of normalizing factor

During the training process, a small normalizing factor was introduced and was found to have a good deal of influence over final accuracy values. A repeat of the above experiment was done without using the normalizing factor to obtain the values in Table 5.4. The images used for training and testing were not histogram equalized in this case. The normalizing value was obtained by finding the trace of the centered image matrix. The centered image matrix was further modified by dividing this value. During the testing phase, centered test images were normalized further using this value and then projected onto the nonlinear subspace. The values shown in Table 5.2 were obtained using this normalizing factor.

Table 5.4: Effect of normalizing value.

Method	Accuracy(%/#)
Normalizing factor	94.54(156/165)
Without factor	70.9(117/165)

5.2.2 Importance of training parameters

It was also found that blindly applying the above factors to images resulted in decreased accuracy. The above tests were conducted such that all factors were used consistently during calculation of the features and during the projection phase. Now we set up the tests such that the normalizing factor is used during only one phase of the procedure. Table 5.5 consolidates results from all the combinations we tested. The best possible result was obtained when the images were histogram equalized and normalized during both phases of the procedure. The worst possible combination was when histogram equalized and normalized images were used for computing the nonlinear components and original images were projected using these components as shown in row 4. Not using the two steps in both phases resulted in comparatively better accuracy as shown in row 5. By creating this artificial variability we show the possibility of incorrect projection if the data lies far away from the modeled curve, leading to possible wrong classification.

Table 5.5: Training parameter variations.

Feature extraction	Projection	Accuracy(%/#)
Histogram equalized/Normalized	Histogram equalized/Normalized	98.1(162/165)
Histogram equalized/Normalized	Not equalized/Normalized	69.69(115/165)
Histogram equalized/Normalized	Histogram equalized/Not normalized	83.3(137/165)
Histogram equalized/Normalized	Not equalized/Not normalized	41.8(69/165)
Not equalized/Not normalized	Not equalized/Not normalized	70.9(117/165)

5.3 CMU expression variant face database

A set of images from the CMU expression variant face database [68] is considered to illustrate the efficiency of the proposed method for expression invariant face recognition. The database consists of face images of 13 different individuals with 75 different images for each individual. The images are taken under constant illumination condition and from the same pose and vary only in facial expressions. The first two components of the nonlinear subspace approach for 3 of the individuals from the database are plotted against each other and shown in Figure 5.5. While we can observe a definite arrangement, the patterns are not well separated in this approach. Figure 5.6 shows the result for the nonlinear discriminant approach. We can observe that by using this method, we are able to separate individuals into well defined and easily classifiable clusters.

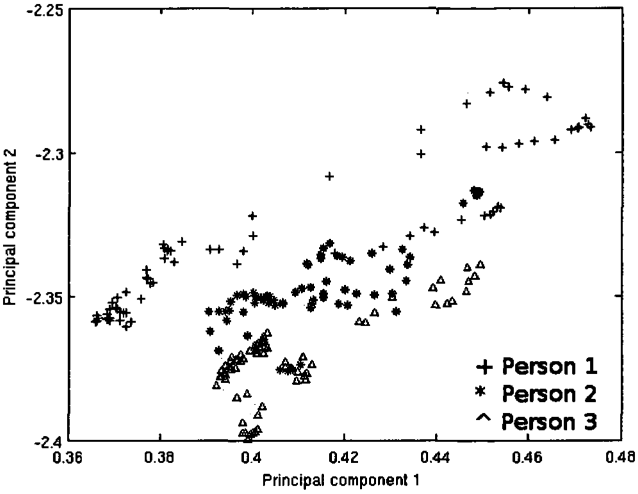


Figure 5.5: Projection of multiple persons to first two nonlinear components.

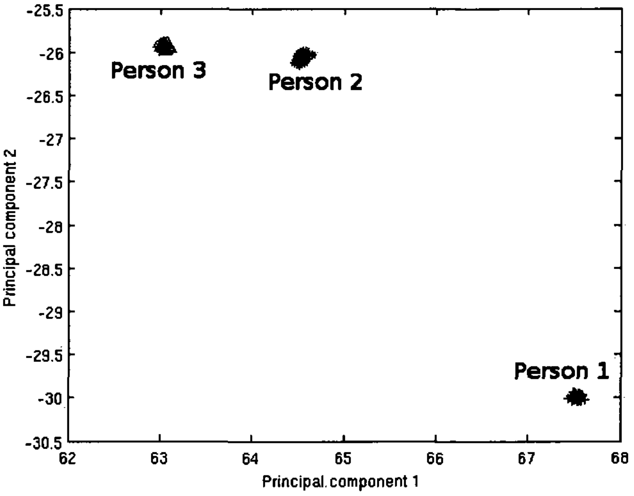


Figure 5.6: Projection of multiple persons to first two nonlinear discriminant components.

Within each separate cluster in the 2D feature space, face images show a definite arrangement with varying expressions as is observable in Figure 5.5. A similar approach can be found in [3] where a nonlinear manifold approach is used to model face expressions. Weight values from the first two principal components are plotted against each other to obtain a distinct cluster for each expression as illustrated in Figure 5.7. Face recognition accuracy for this database is tested by varying the number of dimensions and checking the number of correct classification in each dimension. The whole dataset was divided into two groups: the training set containing 25 images, and the test set containing the remaining 50 images. Figure 5.8 compares the accuracy values across varying dimensions of the feature space for PCA and the proposed discriminant approach.

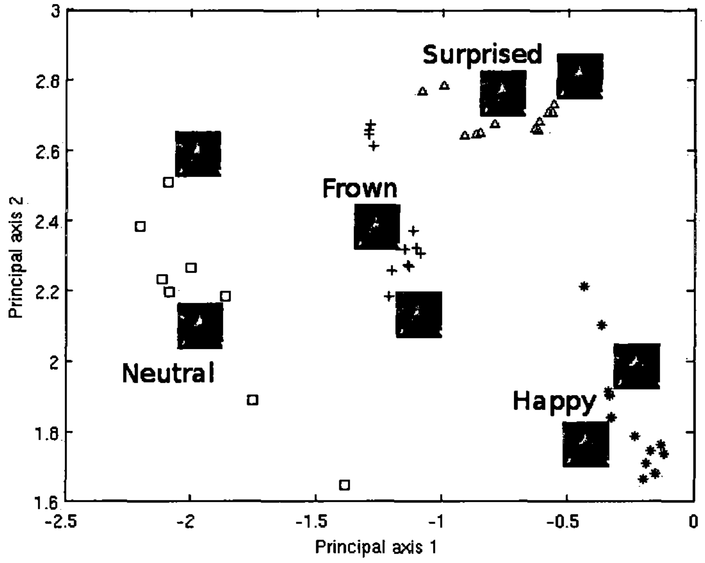


Figure 5.7: Illustration of face expression cluster formation.

5.4 FERET database

FERET was a general evaluation designed to compare performance of algorithms on the FERET database [65], [66]. Different categories of images are provided to test the robustness of algorithms. Categories differed in lighting changes, people wearing

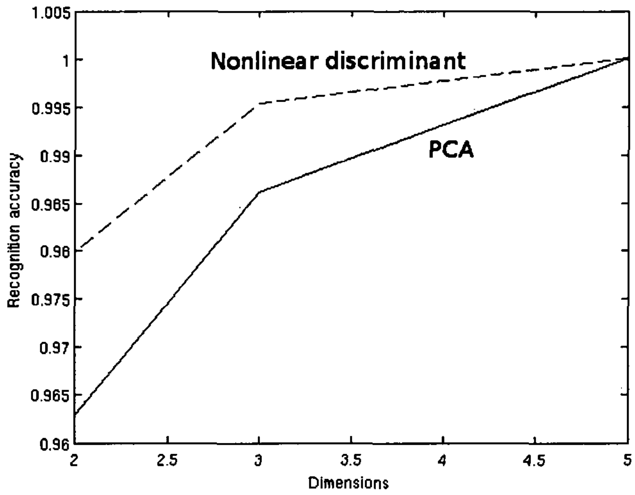


Figure 5.8: Recognition accuracy on CMU expression variant face database.

glasses and time and date of image acquisition. Algorithms that can be applied to this database are broadly divided into two categories: fully automatic algorithms that incorporate face localization and normalization and partially automatic algorithms that use eye center data provided with the database to perform localization and normalization. Performance results for both cases are distinguished and compared separately. The proposed algorithm falls into the latter category. We use the eye center data to localize the face image and each image is resized to a size of 64 × 64 pixels. The main issue we had with this database is the absence of multiple images for training purposes. The gallery and probe consists of only one image of each individual. Hence, a generalized set from the Yale face database described above was used to train the feature components. These components are now used to project images on to the feature space.

5.4.1 FERET testing strategy

The algorithm is given two sets of images: the *target* set and the *query* set. Images in the two sets are distinct. The target set consists of a set of known facial images. Images in the query set consist of unknown facial images to be identified. For each image q_i in the query set Q, we compute a similarity $s_i(k)$ between q_i and each image t_k in the target set T. For each q_i, target images t_k are sorted by the similarity scores $s(\cdot)$.

The following method, as given in [65], is used to compute the similarity score. Let $P = \{p_1, \cdots, p_N\}$ denote a probe set and N be the number of images (individuals). This is matched against a gallery $G = \{g_1, \cdots g_M\}$ by comparing the similarity scores $s_i(\cdot)$ such that $p_i \in P$ and $g_k \in G$. A smaller similarity score implies a closer match. Let $id(i)$ denote the index of the gallery image of the person in probe p_i. A probe p_i is correctly identified if $s_i(id(i))$ is the smallest score for $g_k \in G$. A probe p_i is in the top n if $s_i(id(i))$ is one of the n^{th} smallest scores of $s_i(\cdot)$ for gallery G. Now if R_n denotes the number of probes in the top n, the rank (or cumulative match score) of the probe set is defined as,

$$rank = \frac{R_n}{N} \tag{5.1}$$

5.4.2 FERET face datasets

Three different set probes are used here for testing purposes. All three use the same gallery of 1196 individuals for training. For the first test, the target and query sets contain frontal images (**fa** and **fb**) taken in the same session. The **FB** probe set contains the **fb** images that are taken the same day under the same illumination conditions as the gallery images (**fa**). The second probe category contains all duplicate frontal images in the FERET database for the gallery images. This is referred to as the duplicate set of probe (**duplicate I**) and contains a set of 722 images. The third category contains 234 **fc** probe images. These are images taken the same day as the **fa** images but with a different camera and lighting. A sample image from each of the datasets is shown in Figure 5.9.

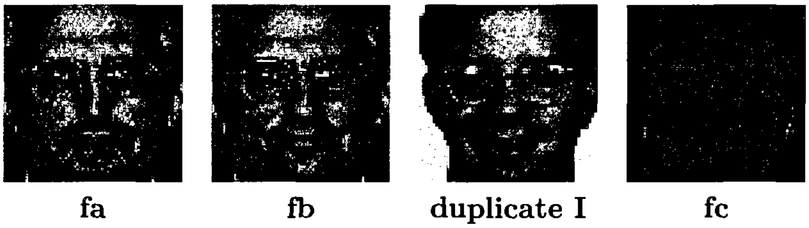


Figure 5.9: FERET images.

The result from the **FB** probe set is given in Figure 5.10. The nonlinear manifold method with modularity performed the best. The poor performance of the nonlinear discriminant method can be attributed to the absence of well defined class information in this case. The phase congruent approach did not perform as well as expected, scoring only marginally better than the baseline PCA method. The score for the modular approach is better than the baseline PCA and compares well with most methods listed in [65], but in cases where the classification strategy varies, like the probabilistic subspaces instead of the nearest neighbor approach, the results are better than the proposed method. Figure 5.11 shows the result for the duplicate I probe set and Figure 5.12 shows results for the **fc** probe set with baseline PCA method for reference.

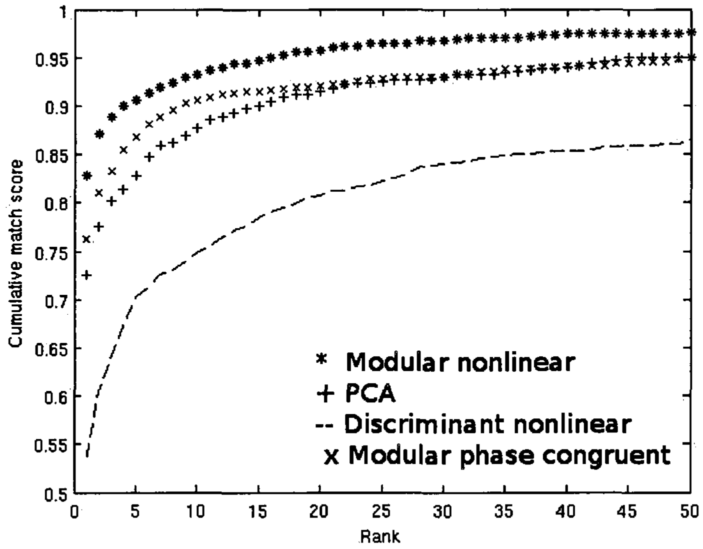


Figure 5.10: FERET result for **FB** probe set.

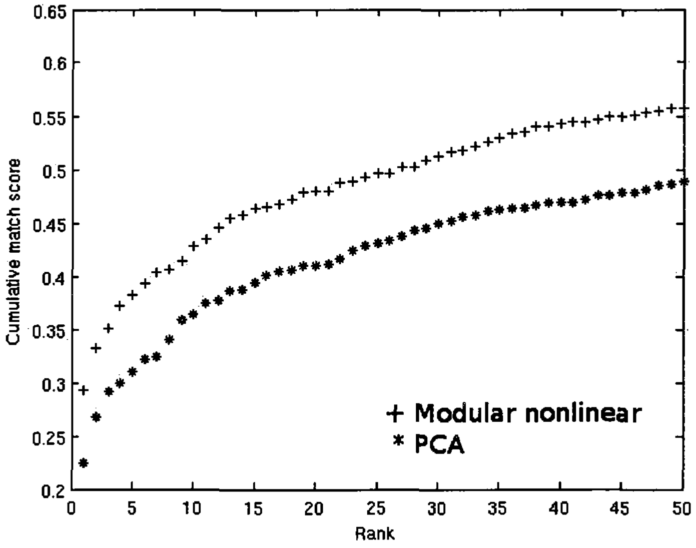


Figure 5.11: FERET result for duplicate I set.

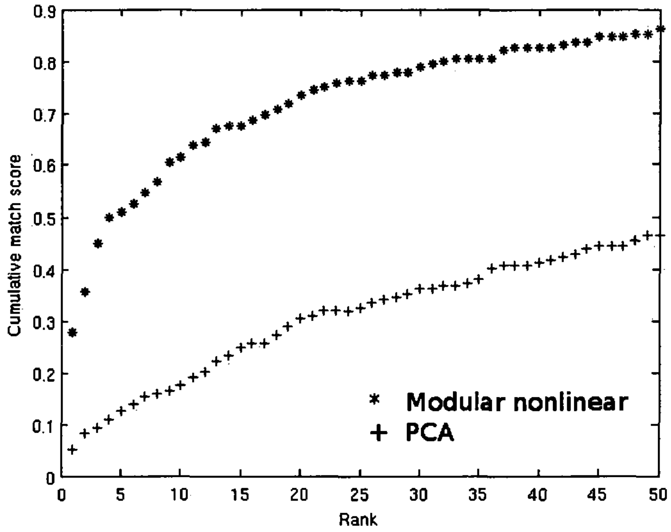


Figure 5.12: FERET result for fa-fc set.

6 Conclusion and Future Work

A nonlinear subspace projection methodology for extracting key feature components for pattern classification was presented in this dissertation. Data covariance in a class was represented as a nonlinear function by using a second order polynomial curve equation. Least squares estimation was used on centered data values to find covariance components in each order. Eigen decomposition of these covariance matrices gave the required Principal Components. The original data pattern was projected to the feature space by multiplying with these features. By choosing only a select number of eigenvectors (those corresponding to the largest eigenvalues) corresponding to each order, each data pattern was projected on to a low dimensional feature space.

The proposed method was used to extract face image features and the extracted features were classified using a nearest neighbor strategy. In a multi-class problem such as face recognition, a polynomial curve was modeled for each class. A mean curve was then obtained by averaging curves from all the classes. Eigen decomposition was then applied to these mean covariance matrices (multiple orders) to obtain the required features. This strategy allowed us to project patterns onto one global feature space to make meaningful comparisons.

A modular approach was followed to improve the accuracy of our system. This approach used multiple curves to represent each data pattern, possibly allowing for simpler curves. To do this, each image was divided into sub-images and nonlinear principal components were extracted for each of these sub-images. These values were then concatenated to form a signature vector. The approach was better able to represent local information and thus lead to better accuracy values.

The proposed method was further extended to improve face recognition accuracy by incorporating class information thus leading to better discriminating features. To this end, a Fisher Linear Discriminant (FLD) based algorithm was developed by computing within-class and between-class scatter matrices. One issue that we faced was the singularity problem of the within-class scatter matrix due to small number of training samples. To solve this issue, the discriminant method was applied on a reduced feature space obtained by the nonlinear subspace method.

The proposed methods were tested on four different databases. The discriminant approach gave the best results in the Yale, PIE illumination and CMU expression variant face databases. The approach, though, was found to work well only if we have well defined classes as it gave poor results in the FERET database, which provided only one image of each individual in each gallery and probe. The proposed method was found to model variations in illumination well as shown by the good results from the PIE illumination database. A phase congruent approach was followed to further improve accuracy but was found to perform poorer than the nonlinear modular approach above. One of the issues with nonlinear manifold based methods is the inability of these methods to project data back to the original space and reconstruct the images. We were able to show that the proposed method is able to recover data from the nonlinear feature space.

Table 6.1 gives a comparison of the proposed method with Isomap. Unlike Isomap and most other nonlinear manifold method, the proposed approach is not neighborhood dependent. The proposed approach provides a direct projection to the feature space and data can be reconstructed from the feature space.

Table 6.1: A comparison with Isomap.

Isomap	Proposed approach
Neighborhood size dependent	Not dependent
No direct projection to feature space	Direct projection to feature space
Data cannot be reconstructed	Data can be reconstructed
Model variations in all directions	Model variations along one direction

One of the issues of the proposed method is its inability to model variations along multiple directions. A more local approach that divides data into smaller clusters could solve this problem. Research work is progressing to divide datasets of similar patterns into neighborhoods that will enable feature space transformation with reduced complexity. The trick here is to combine all the multiple curves into one single region to enable direct projection of new data. Another approach that might provide improved accuracy for the proposed approach is to use Gabor wavelets based frequency representation. Gabor wavelet based image representation is widely applied in face recognition [69, 70, 71]. This particular representation of images is similar to those of the human visual system and found to provide improved accuracy over conventional representation of images. Another topic to be investigated is to try multiple classification strategies to find one that best suite our algorithm. Current work is also progressing to implement the subspace strategy with other nonlinear functions and to identify the optimum nonlinear function to be used from an object classification perspective.

REFERENCES

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.

[2] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 4–37, January 2000.

[3] M. J. Seow and V. K. Asari, "Nonlinear line attractor network for face expression association," in *Proceedings of the International Conference on Artificial Neural Networks In Engineering - ANNIE 2005*, pp. 553–558, 2005.

[4] I. T. Jolliffe, *Principal Components Analysis.* New York: Springer-Verlag, 1986.

[5] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[6] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[7] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. Mersereau, "Eigenface-based super-resolution for face recognition," in *Proceedings of International Conference on Image Processing*, vol. 2, pp. 845–848, 2002.

[8] E. Yoruk, H. Dutagaci, and B. Sankur, "Hand biometrics," *Image and Vision Computing*, vol. 24, pp. 483–497, May 2006.

[9] J. Ahlberg, “Facial feature extraction using eigenspaces and deformable graphs,” in *Proceedings of the International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging*, pp. 8–11, 1999.

[10] J. B. Tenenbaum, V. de. Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319 – 2323, 2000.

[11] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323 – 2326, 2000.

[12] M. H. C. Law and A. K. Jain, “Incremental nonlinear dimensionality reduction by manifold learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 377–391, 2006.

[13] O. Samko, A. D. Marshall, and P. L. Rosin, “Selection of the optimal parameter value for the isomap algorithm,” *Pattern Recognition Letters*, vol. 27, no. 9, pp. 968–979, 2006.

[14] M.-J. Seow and V. Asari, “Recurrent neural network as a linear attractor for pattern association,” *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 246–250, 2006.

[15] M.-J. Seow and V. K. Asari, “Recurrent network as a nonlinear line attractor for skin color association,” *Lecture Notes in Computer Science*, vol. 3173/2004, pp. 870–875, 2004.

[16] H. Seung and D. Lee, "The manifold ways of perception," *Science*, vol. 290, pp. 2268–2269, December 2000.

[17] P. Chellapa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[18] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, 1993.

[19] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America*, vol. 4, no. 3, pp. 519–524, 1987.

[20] A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," *International Journal of Computer Vision*, vol. 18, pp. 233–254, June 1996.

[21] M.-H. Yang, D. J. Kriegman, S. Member, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34–58, 2002.

[22] A. H. Gee and R. Cipolla, "Determining the gaze of faces in images," *Image and Vision Computing*, vol. 12, pp. 639–647, 1994.

[23] T. Horprasert, Y. Yacoob, and L. S. Davis, "Computing 3-d head orientation from a monocular image sequence," in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pp. 242–247, 1996.

[24] B. Shashua, D. Beymer, A. Shashua, and T. Poggio, "Example based image analysis and synthesis," in *MIT AI Memo*, pp. 1–21, 1993.

[25] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions On Pattern Analysis and Machine intelligence*, vol. 20, pp. 23–38, 1996.

[26] P. Viola and M. Jones, "Robust real-time object detection," in *International Journal of Computer Vision*, vol. 1, pp. 511–518, 2001.

[27] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 780–788, 2002.

[28] I. Borg and P. J. F. Groenen, *Modern Multidimensional Scaling : Theory and Applications.* Springer, 2005.

[29] R. Gottumukkal and V. K. Asari, "An improved face recognition technique based on modular pca approach," *Pattern Recognition Letters*, vol. 25, no. 4, pp. 429–436, 2004.

[30] P. Sankaran and V. Asari, "A multi-view approach on modular pca for illumination and pose invariant face recognition," *in Proceedings of the IEEE Applied Imagery and Pattern Recognition*, pp. 165–170, 2004.

[31] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computing*, vol. 10, no. 5, pp. 1299–1319, 1998.

[32] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, pp. 181–201, 2001.

[33] R. Courant and D. Hilbert, *Methods of Mathematical Physics, vol 1.* New York: Interscience, 1953.

[34] K. Kim, K. Jung, and H. Kim, "Face recognition using kernel principal component analysis," *Signal Processing Letters*, vol. 9, pp. 40–42, February 2002.

[35] M.-H. Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 215–220, 2002.

[36] S. K. Zhou, "Intra-personal kernel space for face recognition," in *Proceedings of the IEEE International Automatic Face and Gesture Recognition*, pp. 235–240, 2004.

[37] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX*, pp. 41–48, 1999.

[38] C. K. Williams, "On a connection between kernel pca and metric multidimensional scaling," in *Advances in Neural Information Processing Systems 13*, pp. 675–681, 2001.

[39] J. Choi and J. Yi, "Low-dimensional facial image representation using fld and mds," in *Perceptual and Pattern Recognition: Advances in Intelligent Computing*, pp. 223–232, 2005.

[40] H. Choi and S. Choi, "Robust kernel isomap," *Pattern Recognition*, vol. 40, no. 3, pp. 853 – 862, 2007.

[41] X. Li and L. Shu, "Kernel based nonlinear dimensionality reduction for microarray gene expression data analysis," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7644 – 7650, 2009.

[42] Y. H. Herlin, H. M. Yahia, E. G. Huot, and I. L. Herlin, "Geodesic distance evolution of surfaces: A new method for matching surfaces," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 663–668, 2000.

[43] R. W. Floyd, "Algorithm 97: Shortest path," *Communications of the ACM*, vol. 5, no. 6, p. 345, 1962.

[44] M. Balasubramanian and E. L. Schwartz, "The isomap algorithm and topological stability," *Science*, vol. 295, p. 7, January 2002.

[45] M.-H. Yang, "Extended isomap for pattern classification," in *Eighteenth national conference on Artificial intelligence*, (Menlo Park, CA, USA), pp. 224–229, American Association for Artificial Intelligence, 2002.

[46] M.-H. Yang, "Extended isomap for classification," in *ICPR '02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 3*, p. 30615, 2002.

[47] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[48] J. Yang and J.-Y. Yang, "Why can lda be performed in pca transformed space?," *Pattern Recognition*, vol. 36, no. 2, pp. 563 – 566, 2003.

[49] J. Choi and J. Yi, "A two-stage dimensional reduction approach to low-dimensional representation of facial images," in *Biometric Authentication: Lecture Notes in Computer Science*, pp. 1–9, 2004.

[50] J. A. Lee, A. Lendasse, and M. Verleysen, "Nonlinear projection with curvilinear distances: Isomap versus curvilinear distance analysis," *Neurocomputing*, vol. 57, pp. 49 – 76, 2004.

[51] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, pp. 1373–1396, 2002.

[52] X. He, S. Yan, Y. Hu, H.-J. Zhang, and P. Niyogi, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[53] V. D. Silva and J. B. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction," in *Advances in Neural Information Processing Systems*, vol. 15, pp. 705–712, 2003.

[54] M. Brucher, C. Heinrich, F. Heitz, and J.-P. Armspach, "A metric multidimensional scaling-based nonlinear manifold learning approach for unsupervised data reduction," *EURASIP Journal on Advances in Signal Processing*, vol. 2008 (2008), p. 12, 2008.

[55] B. Kégl, A. Krzyzak, T. Linder, and K. Zeger, "Learning and design of principal curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, pp. 281–297, 2000.

[56] S. Pal, P. K. Biswas, and A. Abraham, "Face recognition using interpolated bezier curve based representation," *in Proceedings of the International Conference on Information Technology: Coding and Computing*, vol. 1, p. 45, 2004.

[57] P. Sankaran, R. Gottumukkal, and V. K. Asari, "An automated feature localisation algorithm for a feature specific modular approach for face recognition," *International Journal of Intelligent Systems Technologies and Applications*, vol. 2, no. 4, pp. 329–344, 2007.

[58] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images (invited paper)," in *AVBPA '97: Proceedings of the First International Conference on Audio- and Video-Based Biometric Person Authentication*, pp. 127–142, 1997.

[59] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in *FG '98: Proceedings of the Third IEEE International Conference on Face and Gesture Recognition*, p. 336, 1998.

[60] S. Gundimada and K. V. Asari, "A novel neighborhood defined feature selection on phase congruency images for face recognition in the presence of extreme variations," *International Journal of Information Technology*, vol. 3, no. 1, pp. 25–31, 2006.

[61] P. Kovesi, "Edges are not just steps," in *The 5th Asian Conference on Computer Vision*, pp. 23–25, 2002.

[62] P. Kovesi, "Image features from phase congruency," *Videre: Journal of Computer Vision Research*, vol. 1, no. 3, pp. 1–27, 1999.

[63] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *Workshop on Applications in Computer Vision 94*, pp. 138–142, 1994.

[64] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643 – 660, 2001.

[65] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[66] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi, "The feret evaluation methodology for face-recognition algorithms," in *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, p. 137, 1997.

[67] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1615–1618, 2003.

[68] X. Liu, T. Chen, and B. V. K. V. Kumar, "Face authentication for multiple subjects using eigenflow," *Pattern Recognition*, vol. 36, no. 2, pp. 313 – 328, 2003.

[69] N. Gudur and V. Asari, "Gabor wavelet based modular pca approach for expression and illumination invariant face recognition," in *AIPR '06: Proceedings of the 35th IEEE Applied Imagery and Pattern Recognition*, p. 13, 2006.

[70] B. Li and R. Chellappa, "Gabor attributes tracking for face verification," in *International Conference on Image Processing*, pp. 45–48, 2000.

[71] W. Choi, S. Tse, K. Wong, and K. Lam, "Simplified gabor wavelets for human face recognition," *Pattern Recognition*, vol. 41, pp. 1186–1199, March 2008.

List of publications

Journals

1. P. Sankaran, R. Gottumukkal, and K. V. Asari, "An automated feature localization algorithm for a feature specific modular approach for face recognition," *International Journal of Intelligent Systems Technologies and Applications*, vol. 2, no.4, pp. 329-344, 2007.

2. P. Sankaran, S.Gundimada , D.Valaparla , and K. V. Asari, "Eye center location based on estimated head pose," *Multimedia Cyberscape Journal: Special Issue on Multimedia Data Processing and Compression*, vol. 2, no. 4, 2005.

Conferences

3. P. Sankaran and K. V. Asari, "Nonlinear Manifold Discriminant Embedding for Face Recognition," *IS&T and SPIE conference on Electronic Imaging: Applications of Artificial Neural Networks in Image Processing XII,* San Jose, California, January 2009.

4. P. Sankaran and K. V. Asari, "Non linear Manifold Representation of a Face Image for Classification," *International Conference on Artificial Neural Networks in Engineering-ANNIE*, St.Louis, MO, November 9-12, 2008.

5. N. Unaldi, P. Sankaran, K. V. Asari and Z. Rahman, "Image Enhancement

for Improving Face Detection Under Non-Uniform Lighting Conditions," *ICIP*, San Diego 2008.

6. I. Purohit, P. Sankaran, K. V. Asari and M. A. Karim, "A modular approach on adaptive thresholding for the extraction of mammalian cell regions from bioelectric images in complex lighting environments," *SPIE Defense and Security Symposium: Visual Information Processing XVII*, Orlando, FL, March 16-20, 2008.

7. P. Sankaran and K. V. Asari, "Adaptive thresholding based mammalian cell segmentation for cell-destruction activity verification," *IEEE International Workshop on Applied Imagery and Pattern Recognition, AIPR - 2006*, Washington DC, October 11 - 13, 2006.

8. K. V. Asari, R.Gottumukkal , R. C. Tompkins, and P. Sankaran, "Pose and lighting invariant face detection and recognition," *3rd Face Recognition Grand Challenge Workshop, FRGC 2005*, MITRE, McLean, Virginia, February 16, 2005.

9. P. Sankaran, S.Gundimada , R. C. Tompkins, and K. V. Asari, "Pose angle determination by face, eyes and nose localization," IEEE Workshop on Face Recognition Grand Challenge Experiments , *FRGC 2005 in conjunction with the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2005*, San Diego, CA, June 20 , 25, 2005.

10. P. Sankaran and K. V. Asari, "A multi-view approach on modular PCA

for illumination and pose invariant face recognition," *IEEE International Workshop on Applied Imagery and Pattern Recognition, AIPR - 2004*, Washington DC, USA, ISBN 0-7695-2250-5, pp. 165-170, October 2004.

Published abstracts

11. P. Sankaran, V. K. Asari and M. A. Karim, "Nonlinear Manifolds for Face Recognition", *ODU-NSU-EVMS-VTC Research Exposition Day: Research Expo 2008,* Ted Constant Hall, Norfolk, VA, pp. 27, April 09, 2008.

12. P. Sankaran, H. Ngo, V. Asari and M. Karim, "Robotic nose for detection of improvised explosive devices", *ODU-NSU-EVMS-VTC Research Exposition Day: Research Expo 2007 - 400 Years of Discovery*, Ted Constant Hall, Norfolk, VA, p. 44, April 05, 2007.

13. P. Sankaran, D. E. Bitner, and K. V. Asari, "Explosive detection using chemical vapor sensors," *ODU-NSU-EVMS Research Exposition Day: Research Expo 2006 - Global Challenges and Local Solutions*, Ted Constant Hall, Norfolk, VA, p. 38, April 05, 2006.

14. P. Sankaran, S. Gundimada, R. Gottumukkal, R. C. Tompkins, W. D. Oden and K. V. Asari, "Face preprocessing for face recognition and face recognition in video," *ODU Research Day*, Old Dominion University, Norfolk, April 6, 2005.

VITA

Praveen Sankaran is a PhD student in Electrical and Computer Engineering at Old Dominion University (ODU), Norfolk, Virginia. Praveen received the Bachelors degree in Applied Electronics and Instrumentation from University of Calicut, India, in 2002 and the Masters degree in Electrical Engineering from ODU, in 2005.

Educational background

PhD, Electrical and Computer Engineering, Old Dominion University, December 2009.

Masters, Electrical Engineering, Old Dominion University, August 2005.

Bachelors, Applied Electronics and Instrumentation, University of Calicut, India, August 2002.

Department of study

Department of Electrical and Computer Engineering, 231, Kaufman Hall, Norfolk, VA, 23529.

Research laboratory

Computational Intelligence and Machine Vision Laboratory, Suite 202, 4111, Monarch Way, Norfolk, VA, 23508.