Performance Evaluation of PCA and LDA for Face Recognition

S. K. Hese, M. R. Banwaskar

Department of Electronics & Telecommunication, MGM’s College of Engineering Nanded
Near Airport, Nanded, Maharashtra, India

Abstract — Automated face recognition has become a major field of interest. In this field several facial recognition algorithms have been explored in the past few decades. Progress has been made towards recognition under varying lighting conditions, poses and facial expressions. In a general context, a facial recognition algorithm and its implementation can be considered as a system. The input to the facial recognition system is a two dimensional image, while the system distinguishes the input image as a user’s face from a pre-determined library of faces. Finally, the output is a discerned face image.

This paper discusses different appearance based face recognition techniques. The experimentation includes the use of image preprocessing techniques followed by most popular dimensionality reduction algorithms based on PCA and LDA. Here our aim is to evaluate the performance of face recognition algorithms based on principle component analysis and linear discriminant analysis on small training data set. The result obtained showed that PCA outperforms LDA.

Keywords: Face recognition, pattern recognition, image preprocessing, grayscale conversion, histogram equalization, PCA, LDA.

I. INTRODUCTION

Face is a complex multidimensional structure and needs good computing techniques for recognition. The face is our primary and first focus of attention in social life playing an important role in identity of individual. We can recognize a number of faces learned throughout our lifespan and identify that faces at a glance even after years. There may be variations in faces due to aging and distractions like beard, glasses or change of hairstyles.

Face recognition is an integral part of biometrics. In biometrics basic traits of human is matched to the existing data and depending on result of matching identification of a human being is traced. Facial features are extracted and implemented through algorithms which are efficient and some modifications are done to improve the existing algorithm models. Face recognition algorithms are used in a wide range of applications viz., security control, crime investigation, and entrance control in buildings, access control at automatic teller machines, passport verification, identifying the faces in a given databases [1], [2].

The Eigenface is the first method considered as a successful technique of face recognition. The Eigenface method uses Principal Component Analysis (PCA) to linearly project the image space to a low dimensional feature space [3], [4]. The Fisherfaces is an enhancement of the Eigenface method. The Eigenface method uses PCA for dimensionality reduction, thus, yields projection directions that maximize the total scatter across all classes, i.e. across all images of all faces. The PCA projections are optimal for representation in a low dimensional basis, but they may not be optimal from a discrimination standpoint. Instead, the Fisherfaces method uses Fisher’s Linear Discriminant Analysis (FLDA or LDA) which maximizes the ratio of between-class scatter to that of within-class scatter [5].

At one level, PCA and LDA are very different: LDA is a supervised learning technique that relies on class labels, whereas PCA is an unsupervised technique. LDA has been compared to PCA in several studies [6], [7], [8]. One characteristic of both PCA and LDA is that they produce spatially global feature vectors. In other words, the basis vectors produced by PCA and LDA are non-zero for almost all dimensions, implying this change to a single input pixel, will alter every dimension of its subspace projection.

As we know that any image or face has size n x m pixels which require n.m dimensional space. This space is too large and needs to be reduced for better recognition which is achieved by dimensionality reduction techniques [1]. We have two most popular techniques for these purposes that are principal component analysis (PCA) and linear discriminant analysis (LDA) [7]. For better performance we have implemented these two algorithms with several preprocessing factors such as gray scale conversion and modified histogram equalization before recognition algorithms. The aim of this paper is to study the performance of the PCA and LDA with respect to face recognition rate and dimensionality. Considering for small training data set, we have designed the both algorithms for face recognition. The experiments are based on Yale database. The organization of this paper is done in six sections. Section II describes the preprocessing methods performed on facial images. Section III provides introduction to PCA and its mathematical derivation. Section IV discusses LDA and the related mathematical analysis and results and conclusion are presented in section V & VI respectively.

II. PREPROCESSING OF IMAGES

Image preprocessing techniques represent an essential part of face recognition systems, which has a great
impact on the performance and robustness of the recognition procedure. The main objective of these techniques is to enhance the discriminative information contained in the facial images.

A. Grayscale Conversion

In order to retain as much as information of images, the color images are converted into grayscale images. This is the first step of experiment. As color images (RGB images) are composed of 3 channels to present red, green and blue components in RGB space. Pixels in grayscale images are stored as 8-bit integer to represent color into black and white [9], [10].

B. Histogram Equalization

It is usually done on low contrast images in order to enhance image quality and to improve face recognition performance. It changes the dynamic range (contrast range) of the image and as a result, some important facial features become more visible. An image histogram is a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. In Histogram Equalization the global contrast enhancement is obtained using the cumulative density function of the image as a transfer function [9], [11]. The result is a histogram approximately constant for all gray values. Mathematically histogram equalization can be expressed as:

$$S_k = T(r_k) = \sum_{j=0}^{k} \frac{n_j}{n}$$

Whereas $k = 0, 1, 2, \ldots, L-1$. Here in above equation $n$ is the total number of pixels in an image, $n_j$ is the number of pixels with gray level $r_k$ and $L$ is the total number of gray levels exist in the face image. Fig. 1 depicts the image histogram before equalization and Fig. 2 clearly shows the effect of histogram equalization after equalization processing on image.

III. PRINCIPAL COMPONENT ANALYSIS

Research in automatic face recognition started in the 1960’s. Kirby and Sirovich were among the first to apply principal component analysis (PCA). Turk and Pentland popularized the use of PCA for face recognition. Principle Component Analysis (PCA) is a dimensionality reduction technique that is used for image recognition and compression. It is also known as Karhunen-Loeve transformation (KLT) or eigenspace projection [3], [4].

A. Eigenface approach

As proposed by Turk and Pentland, the system was initialized or trained with the following operations:

1. An initial set of face images were acquired. This was the training set.
2. The Eigenfaces were calculated from the training set. Only M Eigenfaces corresponding to the M largest Eigenvalues were retained. These Eigenfaces spanned the face space which constituted of the training set.
3. The M Eigenface-weights were calculated for each training image by projecting the image onto face space spanned by the Eigenfaces. Each face image then will be represented by M weights- an extremely compact representation. After initialization, the following steps were performed to recognize test images:
4. The set of M weights corresponding to the test image were found by projecting the test image onto each of the Eigenfaces.
5. The test image was determined if it was a face at all by checking whether it was sufficiently close to the face space. This was done by comparing the distance between the test image and the face space to an arbitrary distance threshold.
6. If it was sufficiently close to the face space, compute the distance of the M weights of the test image to the M weights of each face image in the training set. A second arbitrary threshold was put in place to check whether the test image corresponded at all to any known identity in the training set.
7. If the second threshold was overcome, the test image was assigned with the identity of the face image with which it had the smallest distance.
8. For a test image with a previously unknown identity, the system was retrained by adding this image to the training set.

The following figure (Fig. 3) shows some of the examples of eigenfaces.
B. Mathematics of PCA

An image \( I(x, y) \) whose dimensions is \((N \times N)\) can be considered as a vector of dimension \( N^2 \). For example a typical image of size \((512 \times 512)\) will become a vector of dimension \(262,144\). This is equivalent to a point in the 262,144-dimensional space. An ensemble of images would therefore map to a collection of points in this huge space. Images of faces being similar in overall configurations will not be randomly distributed in this huge image space and can be described by a relatively low dimensional subspace. The main idea of PCA is to find the vectors that best account for the distribution of face images within the entire image space [12], [13]. For a database of face images if the selected training set is \( \Gamma_1, \Gamma_2, \ldots, \Gamma_M \), then the average face of the set will be given as follows:

\[
\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n
\]

Each face differs from the average face by the vector

\[
\Phi_i = \Gamma_i - \Psi
\]

The set of large vectors is subject to principle component analysis which seeks a set of \( M \) orthonormal vectors; \( U_{\epsilon} \), which best describes the distribution of the data. The \( k^{th} \) vector \( \Phi_k \), is chosen such that \( \lambda_k \) is a maximum:

\[
\lambda_k = \frac{1}{M} \sum_{n=1}^{M} (U_{\epsilon n}^T \Phi_n)^2
\]

The vectors \( U_{\epsilon} \) and \( \lambda_{\epsilon} \) are the eigenvectors and eigenvalues, respectively, of the covariance matrix as given below.

\[
C = \frac{1}{M} \sum_{n=1}^{M} \Phi_n \Phi_n^T = \Lambda \Lambda^T
\]

Where the matrix \( \Lambda = [\Phi_1, \Phi_2, \ldots, \Phi_M] \). The matrix \( C \) has a dimension of \((N^2 \times N^2)\), and determine \( N^2 \) eigenvectors and eigenvalues, and for typical image sizes, this size would be a very high value. Therefore, we need a computationally feasible method to determine these eigenvectors.

Let \( u \) be eigenvectors of \( \Lambda \Lambda^\top \) such that

\[
A^\top A u_i = \mu_i u_i
\]

Multiplying both sides by \( A \), we have

\[
A A^\top A u_i = \mu_i A u_i
\]

From above equation, we can see that \( A u_i \) are the eigenvectors and \( \mu_i \) of \( C = \Lambda \Lambda^\top \). Following this analysis we can construct the \((M \times M)\) matrix as:

\[
L = \Lambda \Lambda^\top
\]

Where \( L_{\epsilon m} = \Phi_m^\top \Phi_n \) and the \( M \) eigenvectors, \( u_i \) of \( L \) are found. These vectors determine linear combinations of the \( M \) training set face images to form the eigenfaces \( U_{\epsilon} \):

\[
U_{\epsilon i} = \sum_{n=1}^{M} v_{\epsilon n} \Phi_n, \quad i = 1, \ldots, M
\]

With this analysis calculations are greatly reduced, from the order of the number of pixels in the images \((N^2)\) to order of the number of images \((M)\) in the training set.

C. Eigenface Classification of Face Images

Eigenface images calculated from the eigenvectors of principal matrix \( L \) span a basis set with which the face images can be described. A new face image \((\Gamma)\) is transformed into its eigenfaces components (projected into face space) by a simple operation:

\[
w_{\epsilon} = U_{\epsilon i}^\top (\Gamma - \Psi), \quad k = 1, 2, \ldots, M'
\]

Weights form a vector \( \Omega^\top = (w_1, w_2, \ldots, w_M) \) that describes the contribution of each eigenface in representing the input face image. The simplest method for determining which face class provides the best description of an input class \( \Omega_k \) that minimizes the Euclidian distance of \( e_k \)

\[
e_k = ||\Omega - \Omega_k||^2
\]

Where \( \Omega_k \) is vector describing the \( k^{th} \) face class. A face is classified as belonging to class \( k \) when the minimum \( e_k \) is below some chosen threshold \( \theta_k \) which defines maximum allowable distance from face space. Otherwise the face is classified as “unknown” and optionally used to create a new class.

IV. LINEAR DISCRIMINANT ANALYSIS

Linear Discriminant analysis or Fisherfaces method applies the fisher's linear discriminant criterion to overcome the limitations of eigenfaces method, which tries to maximize the ratio of determinant of between classes to the determinant of the within-class scatter matrix of the projected samples. Grouping images of the same class, while separating the images of different classes take place due to the fisher discriminant. Projection of face images on fisher space converts its dimension from \( N^2 \)-dimensional space to \( C \) dimensional space (where \( C \) is the number of classes of images). For example, two sets of points are considered in \( 2 \)-dimensional space projected onto a single line hence depending on the direction the points are either mixed (Fig. 4a) or separated (Fig. 4b). The fisher discriminate to find the line which best separates the points i.e., in order to identify the input test image, the comparison of the projected test image with each training image takes place after which the test image as the closest training image can be identified [5], [6], [14].

Along with eigenspace projection, the training images are also being projected into a subspace. The test images being projected at the same subspace can be identified using a similarity measure and the only difference is the way in the subspace calculations take place. The PCA method is used to extract features which represents face image and the LDA method discriminates different face classes in order to find the subspace.
The within class scatter matrix also named as intra-personal shows the variations due to the different lighting and face expression in the appearance of the same individual whereas the between-class scatter matrix i.e., the extra-personal represents the variations due to identity differences. Henceforth, by applying the above mentioned methods, the projection directions on one hand maximize the distance between the face images of different classes, while on the other hand it minimizes the distance between the face images of same class i.e., it can also be said that by maximizing the between-class scatter matrix $S_b$, while minimizing the within-class scatter matrix $S_w$ takes place in projective subspace [13], [15]. Fig. 5a and 5b represents the good and bad class separation.

The definition of the within class scatter matrix $S_w$ and the between-class scatter matrix $S_b$ can be defined as

$$S_w = \sum_{j=1}^{C} \sum_{i=1}^{N_j} (\Gamma_i^j - \mu_j)(\Gamma_i^j - \mu_j)^T$$

Where $\Gamma_i^j$ the ith sample of class $j$, $\mu_j$ is the mean of class $j$. $C$ is the number of classes, $N_j$ is the number of samples in class $j$.

$$S_b = \sum_{j=1}^{C} (\mu_j - \mu)(\mu_j - \mu)^T$$

Here $\mu$ represent the mean of the clas. The spanning of the subspace of LDA is done by a set of vectors $w = [w_1, w_2, ..., w_d]$, which needs to satisfy

$$w = \arg \max \frac{w^T S_b w}{w^T S_w w}$$

The within class scatter matrix shows how the distribution of the face image takes place closely within classes, whereas the between class scatter matrix describes how the separation of classes take place from each other. During the projection of face images in the discriminant vectors $W$, it is seen that the face images should be distributed closely within classes, likewise, should be separated between classes to the maximum, i.e. the discriminant vector is responsible for minimizing the denominator and maximizing the numerator. Therefore, the construction of $W$ takes place with the help of eigenvectors of $S_w^{-1} S_b$. There are various other methods in order to solve the problem of LDA for example the pseudo inverse method, the subspace method and the null space method.

The LDA approach being similar to the eigenface method uses projection of training images into subspace. The test images are projected into the same subspace and identified using a similarity measure. The only difference is the method of calculating the subspace characterizing the face space. The face which has the minimum distance with the test face image is labeled with the identity of that image. The minimum distance can be calculated using the Euclidian distance method as given in earlier Equation.

The following flowchart (Fig. 6) gives the outline of the present work.

![Flowchart for PCA and LDA approach for face recognition](image-url)
V. EXPERIMENTAL RESULTS

We have used faces from Yale database for present work. MATLAB 10 is used to carry out this research work. First of all, the image preprocessing steps are carried out on the images for improving performance of algorithms. Then by applying principle component analysis and linear discriminant analysis, face recognition is done. Further the performance of PCA and LDA based algorithms was evaluated with respect to face recognition rate and dimensionality. Fig. 7 shows some of the faces from Yale database.

Fig. 7. Faces from Yale data base

After applying PCA and LDA, the recognition process gave the closest matching face from training database for the given test image (as shown in Fig. 8).

Fig. 8. Recognized face by PCA and LDA based algorithms

Then the performance curve of PCA and LDA was plotted as given in Fig. 9. The graph (Fig. 9) clearly indicates that generally, PCA outperforms LDA for small training database and sometimes it may give even 100% recognition rate at particular dimensionality.

VI. CONCLUSION

In this study, we have applied two most popular appearance based face recognition methods i.e. PCA and LDA along with image preprocessing factors on Yale database. The Euclidean Distance based classifiers were used for both methods of face recognition systems. The results obtained shows that PCA outperforms LDA for small training database. As shown in Fig. 9, therefore, it may be concluded that for some value of dimensionality, PCA gives higher recognition rate than LDA. In future we want to extend our work on large face database and compare PCA and LDA for their performances. The work further may be extended for another approach of face recognition like Independent Component Analysis (ICA).

REFERENCES

S. K. Hese received her B.E. degree in Electronics & Telecommunication from M.S. Bidve College of Engineering Latur in 2005 and pursuing M.E. in Electronics & Telecommunication from M.G.M. College of Engineering Nanded.

M. R. Banwaskar is pursuing doctorate degree in Computer Science and Engineering under the Faculty of Engineering and Technology, Swami Ramanand Teerth Marathwada University Nanded. She is associated with M.G.M. College of Engineering Nanded as HOD, Department of Electronics and Telecommunication.