Proceedings of the International Conference on Computer and Information Science and Technology
Ottawa, Ontario, Canada, May 11 – 12, 2015
Paper No. 126

Subspace LDA Methods for Solving the Small Sample Size Problem in Face Recognition

Ching-Ting Huang, Chaur-Chin Chen
Department of Computer Science/National Tsing Hua University
101 KwanFu Rd., Sec. 2, Hsinchu, Taiwan
j60626j@gmail.com;cchen@cs.nthu.edu.tw

Abstract– In face recognition, LDA often encounters the so-called small sample size (SSS) problem, also known as curse of dimensionality. This problem occurs when the dimensionality of the data is quite large in comparison to the number of available training images. One of the approaches for handling this situation is the subspace LDA. It is a two-stage framework: it first uses PCA-based method for dimensionality reduction, and then the LDA-based method is applied for classification. This paper investigates four popular subspace LDA methods: Fisherface, Complete PCA plus LDA, IDAface, and BDPCA plus LDA and compare their effectiveness when handling the SSS problem in face recognition. Experimental results tested on three publically available face databases: JAFFE, ORL, and FEI, show that LDA without reducing image size by PCA projection is the worst and BDPCA plus LDA performs better than the other methods for a huge size of database.

Keywords**:** Face Recognition, Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA)

1. Introduction

Face recognition has recently received significant attention and been an important issue in pattern recognition and image analysis over the last few decades as shown by Samaria et al. (1994), Zhao et al. (2003), and Li et al. (2011). The ultimate goal in face recognition is to develop a computer-based automated system which works reliably under unconstrained conditions, runs quickly and requires minimum training data. However, it is still an on-going research area.

Face recognition methods are of two types, the geometry-based approach and the appearance-based approach. The geometry-based approach aims to locate distinctive features such as eyes, nose, mouth, and chin. Properties of the features and relations such as distances and angles between the features are used as descriptors for face recognition. The appearance-based approach operates directly on the intensities of pixels within the face images and processes an image as a two-dimensional holistic pattern. It extracts features in a subspace derived from the training images. Compared to the geometry-based approach, the appearance-based approach performs more robustly under situations like noise, blurring, andthe variations in illumination, facial expressions and occlusion etc. in images.

During the last few decades, many appearance-based approaches have been developed based on Principal Component Analysis (PCA), called Eigenface, such as Sirovich et al. (1987), Kirby et al. (1990), Turk et al. (1991), and Linear Discriminant Analysis (LDA) by Fisher et al. (1936), called Fisherface such as Belhumeur et al. (1997), Yu et al. (2001), Yang et al. (2003), Zhao et al. (2003), and their variants.

LDA usually outperforms PCA for classification tasks since LDA takes class information into account while PCA does not. However, in face recognition, LDA often faces the so-called small sample size (SSS) problem due to the relatively small number of training images per individual compared to the dimensionality of the image space, and would result in the singularity of the within-class scatter matrix.

A number of approaches have been proposed to address the SSS problem. One of the most successful approaches is the subspace LDA, which is a two-phase framework: first projecting the original images into a subspace, where dimensionality is reduced; next preceding LDA-based methods on the lower dimensional subspace. This paper reports the experimental results on 3 databases of four LDA-based methods: Fisherface (PCA plus LDA) by Belhumeur et al.(1997), Complete PCA plus LDA by Yang et al. (2003), an improved discriminate analysis (PCA plus IDA) as shown by Zhuang et al.(2007), and the Bi-Directional PCA plus LDA (BDPCA plus LDA) as shown by Zuo et al.(2006).

The rest of this paper is to give the mathematical foundation of PCA and LDA, review the four subspace LDA methods and show the results of comparison on three databases followed by a conclusion.

2. Background Review

2. 1. Notations

Let $\{X_1, X_2, \dots, X_N\}$ be a training set of N gray level face images of R rows and C columns, and $X_n(r,c) \in \{0,1,\dots,255\}$, $n = 1,2,\dots,N$, $r = 0,1,\dots,R-1$, $c = 0,1,\dots,C-1$. Let $\bar{X} = \frac{1}{N}\sum_{n=1}^{N} X_n \in \mathbb{R}^{R\times C}$ be the mean image of all samples.

Let $\mathbf{x}_\mathrm{n} \in \mathbb{R}^{RC}$ be a column-vector representation of X_n such that $\mathbf{x}_\mathbf{n}(r*R+c) = X_n(r,c)$, $n = 1,2,\dots,N$, $r = 0,1,\dots,R-1$, $c = 0,1,\dots,C-1$. Assume that each face image belongs to one of K classes, and there are N_k face images in class k, of which the mean image is $\bar{\mathbf{x}}_\mathbf{k} = \frac{1}{N_k}\sum_{\mathbf{x}\in class\ k} \mathbf{x} \in \mathbb{R}^{RC}$, $k = 1,2,\ \dots, K$. Let $\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_\mathbf{n} \in \mathbb{R}^{RC}$ be the mean image of all samples in a column vector form.

2. 2. Scatter Matrices

The total scatter matrix $S_T \in \mathbb{R}^{RC\times RC}$, within-class scatter matrix $S_W \in \mathbb{R}^{RC\times RC}$ and between-class scatter matrix $S_B \in \mathbb{R}^{RC\times RC}$can be defined as

$$S_T = \sum_{n=1}^{N} (\mathbf{x}_\mathbf{n} - \bar{\mathbf{x}})(\mathbf{x}_\mathbf{n} - \bar{\mathbf{x}})^t \tag{1}$$

$$S_W = \sum_{k=1}^{K} \sum_{\mathbf{x}\in class\ k} (\mathbf{x} - \bar{\mathbf{x}}_\mathbf{k})(\mathbf{x} - \bar{\mathbf{x}}_\mathbf{k})^t \tag{2}$$

$$S_B = \sum_{k=1}^{K} N_k (\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^t \tag{3}$$

It's easy to verify that $S_T = S_W + S_B$, and S_W and $S_B$are nonnegative definite. The rank of $S_T \in \mathbb{R}^{RC\times RC}$ is the dimensionality of the range space of S_T, which is denoted as $rank(S_T)$ and equal to the number of nonzero eigenvalues of S_T. If $rank(S_T) = RC$, S_T is called full-rank and it is invertible; if $rank(S_T) < RC$, S_T is singular and have zero eigenvalues. Note that $rank(S_T) \le N-1$, $rank(S_W) \le N-K$ and $rank(S_B) \le K-1$.

2. 3. Principal Component Analysis (PCA)

PCA aims to find a set of projection vectors that map the original RC-dimensional image space into a D_{PCA}-dimensional feature space ($D_{PCA} \ll RC$) such that the projection of the data onto the first projection vector has the largest scatter, and the projection onto the second projection vector, which is orthogonal to the first one, has the second largest scatter, and so on. So PCA attempts to find a set of orthonormal eigenvectors, $\{\mathbf{w_1}, \mathbf{w_2},\ \dots,\ \mathbf{w_{D_{PCA}}}\}$ of S_T, to form the projection matrix $W_{PCA} = [\mathbf{w_1}, \mathbf{w_2},\ \dots,\ \mathbf{w_{D_{PCA}}}]$ corresponding to the D_{PCA} largest eigenvalues, where D_{PCA} is considered as the smallest number such that $\frac{\sum_{d=1}^{D_{PCA}} \lambda_d}{\sum_{d=1}^{RC} \lambda_d} > 85\%$. The new feature vector $\mathbf{y_n} \in \mathbb{R}^{D_{PCA}}$ can be obtained by

$$\mathbf{y}_n = W_{PCA}^t \boldsymbol{x}_n,\ n = 1,2,\dots,N \tag{4}$$

2. 4. Linear Discriminant Analysis (LDA)

LDA as shown by Fisher et al. (1936) is a linear dimensionality reduction technique used for face recognition [Belh1997]. It aims to find the optimal set of discriminant vectors that maps the original RC-dimensional image space into a D_{LDA}-dimensional feature space ($D_{LDA} \ll RC$) such that images from different classes are more separated and images of the same class are more compact. LDA then aims to find a set of orthonormal projection vectors such that the first projection vector maximizes the criterion J_0 defined as

$$J_0(\mathbf{w}) = \frac{\mathbf{w}^t S_B \mathbf{w}}{\mathbf{w}^t S_W \mathbf{w}} \tag{5}$$

The goal of LDA becomes to find a projection matrix $W_{LDA} = [\mathbf{w_1}, \mathbf{w_2}, \ \ldots, \ \mathbf{w_{D_{LDA}}}]$, where

$$S_B \mathbf{w_d} = \lambda_d S_W \mathbf{w_d}, \ d = 1, 2, \ldots, D_{LDA} \tag{6}$$

D_{LDA} is considered as the smallest number such that $\frac{\sum_{d=1}^{D_{LDA}} \lambda_d}{\sum_{d=1}^{RC} \lambda_d} > 85\%$. The new feature vector $\mathbf{y_n} \in \mathbb{R}^{D_{LDA}}$ can be obtained by

$$\mathbf{y}_\mathrm{n} = W_{LDA}^t \mathbf{x}_\mathrm{n}, \ n = 1, 2, \ldots., N \tag{7}$$

2. 5. Discussion

LDA is usually more suitable for solving the classification task than PCA because LDA attempts to model the difference between the classes while PCA is good for dimensionality reduction but does not take class information into account. The major drawback of applying LDA for the face recognition task is that it may encounter the so-called small sample size (SSS) problem. It is because $S_W \in \mathbb{R}^{RC \times RC}$ and $rank(S_W) \leq N - K$. If $N < RC + K$, then $rank(S_W) < RC$, S_W is singular. For example, if the size of an image is 112×92, then RC = 10304 but it is not practical to collect 10304+Kface images.

3. Subspace LDA Methods

In recent years, many researchers have noticed the SSS problem and tried to tackle it using different approaches. The following four subspace LDA methods begin with the projection with $\mathbf{y}_\mathrm{n} = W_{PCA}^t \mathbf{x}_\mathrm{n}$ by PCA as defined in (4) for the first stage of a two-stage framework. Second, it applies the LDA-based algorithm in the reduced subspace to get the optimal projection matrix which will be discussed as follows.

3. 1. Fisherface

Among all of the subspace LDA methods proposed over the last few decades, the Fisherface proposed by Belhumeur *et al*. (1997) has attracted much attention. The projection matrix is computed according to the notations given in Section 2. We compute $S_B' = W_{PCA}^t S_B W_{PCA} \in \mathbb{R}^{D_{PCA} \times D_{PCA}}$ and the within-class scatter matrix $S_W' = W_{PCA}^t S_W W_{PCA} \in \mathbb{R}^{D_{PCA} \times D_{PCA}}$ in the reduced subspace. Construct the projection matrix $W_{LDA} = [\mathbf{w_1^{LDA}}, \ \mathbf{w_2^{LDA}}, \ \ldots \ , \ \mathbf{w_{D_{LDA}}^{LDA}}]$ by solving the generalized eigenvalue problem:

$$S_B' \mathbf{w}_\mathrm{d}^\mathrm{LDA} = \lambda_d S_W' \mathbf{w}_\mathrm{d}^\mathrm{LDA}, \ d = 1, 2, \ldots, D_{LDA} \tag{8}$$

where $\mathbf{w}_\mathrm{d}^\mathrm{LDA}$denotes the eigenvector corresponding to the d-th largest eigenvalue λ_d. The optimal projection matrix $W_{OPT} \in \mathbb{R}^{RC \times D_{LDA}}$ and a reduced face feature vector $\mathbf{y_n} \in \mathbb{R}^{D_{LDA}}$ are given as follows.

$$W_{OPT}^t = W_{LDA}^t W_{PCA}^t \tag{9}$$

$$\mathbf{y}_n = W_{OPT}^t \mathbf{x}_n,\ n = 1, 2, \dots, N. \tag{10}$$

3. 2. Complete PCA plus LDA

The Complete PCA plus LDA as shown by Yang et al. (2003) was suggested to give a theoretically optimal and more efficient algorithm that can overcome the weaknesses of the Fisherface mentioned above. Let $D_{com} = rank(S_T)$. We first find the orthonormal eigenvectors of S_T to construct the projection matrix $W_{com} = [\mathbf{w}_1^{\mathbf{com}},\ \mathbf{w}_2^{\mathbf{com}},\ \dots, \mathbf{w}_{\mathbf{Dcom}}^{\mathbf{com}}]$, then compute the between-class scatter matrix $S'_B = W_{com}^t S_B W_{com} \in \mathbb{R}^{D_{com} \times D_{com}}$, the within-class scatter matrix $S'_W = W_{com}^t S_W W_{com} \in \mathbb{R}^{D_{com} \times D_{com}}$, and the total scatter matrix $S'_T = W_{com}^t S_T W_{com} \in \mathbb{R}^{D_{com} \times D_{com}}$ in the reduced subspace. Let $D_{iso} = rank(S'_W)$. In the D_{com}-dimensional transformed space, let $\{\mathbf{w}_1^{\mathbf{iso}}, \mathbf{w}_2^{\mathbf{iso}}, \dots, \mathbf{w}_{\mathbf{D_{com}}}^{\mathbf{iso}}\}$ be a set of all orthonormal eigenvectors of S'_W. Denote $\widetilde{W}_{iso} = [\mathbf{w}_1^{\mathbf{iso}}, \dots, \mathbf{w}_{\mathbf{D_{iso}}}^{\mathbf{iso}}]$ and $\widehat{W}_{iso} = [\mathbf{w}_{\mathbf{D_{iso}+1}}^{\mathbf{iso}}, \dots, \mathbf{w}_{\mathbf{D_{com}}}^{\mathbf{iso}}]$. Let $\widehat{D} = rank(S'_B)$.Let $\tilde{S}_B = \widetilde{W}_{iso}^t S'_B \widetilde{W}_{iso} \in \mathbb{R}^{D_{iso} \times D_{iso}}$, $\tilde{S}_W = \widetilde{W}_{iso}^t S'_W \widetilde{W}_{iso} \in \mathbb{R}^{D_{iso} \times D_{iso}}$. Construct the projection matrix$\widetilde{W} = [\widetilde{\mathbf{w}}_1, \widetilde{\mathbf{w}}_2, \dots, \widetilde{\mathbf{w}}_{\widehat{\mathbf{D}}}]$ by solving the generalized eigenvalue problem

$$\tilde{S}_B \widetilde{\mathbf{w}}_{\mathbf{d}} = \lambda_d \tilde{S}_W \widetilde{\mathbf{w}}_{\mathbf{d}},\ d = 1, 2, \dots, \widehat{D} \tag{11}$$

Let $\hat{S}_B = \widehat{W}_{iso}^T S'_B \widehat{W}_{iso} \in \mathbb{R}^{(D_{com} - D_{iso}) \times (D_{com} - D_{iso})}$. Find the orthonormal eigenvectors of $\hat{S}_B$ to construct the orthogonal matrix $\widehat{W} = [\widehat{\mathbf{w}}_1, \widehat{\mathbf{w}}_2, \dots, \widehat{\mathbf{w}}_{\widehat{\mathbf{D}}}]$. The projection matrix $W_{OPT} \in \mathbb{R}^{RC \times 2\widehat{D}}$ and a reduced feature vector $\mathbf{y_n} \in \mathbb{R}^{2\widehat{D}}$ are obtained by

$$W_{OPT} = W_{com} \times [\widehat{W}_{iso} \widehat{W}, \widetilde{W}_{iso} \widetilde{W}] \tag{12}$$

$$\mathbf{y}_n = W_{OPT}^t \mathbf{x}_n,\ n = 1, 2, \dots, N. \tag{13}$$

3. 3. IDAface

The improved discriminate analysis (IDAface) proposed by Zhuang et al. (2007) is a two-stage framework. First, a modified PCA, called PCA with selection (PCA_S), is used for dimensionality reduction. Second, the algorithm uses inverse Fisher discriminant analysis (IFDA) which introduces a new criterion to derive the discriminative information from both range space and null space of within-class scatter matrix. The inverse Fisher discriminant criterion is defined as

$$\boldsymbol{w} = \arg\min_{W} \frac{|W^t S_W W|}{|W^t S_B W|} \tag{14}$$

Let $p = rank(S_T)$. Select the orthonormal eigenvectors of S_T corresponding to p eigenvalues, which satisfy the inequality $\mathbf{w}^t S_B \mathbf{w} > \mathbf{w}^t S_W$ to form the projection matrix $W_{PCA\_S} = [\mathbf{w}_1, \mathbf{w}_2\ \ \dots, \mathbf{w}_{\mathbf{D_{PCA\_S}}}] \in \mathbb{R}^{RC \times D_{PCA\_S}}$, where $D_{PCA\_S} \leq K - 1$.

Calculate the between-class scatter matrix $S'_B \in \mathbb{R}^{D_{PCA\_S} \times D_{PCA\_S}}$ by $S'_B = W_{PCA\_S}^t S_B W_{PCA\_S}$. Let $D_{proj} = rank(S'_B)$. Work out the orthonormal eigenvectors of S'_B to construct the projection matrix $W_{proj} = [\mathbf{w}_1^{\mathbf{proj}}, \mathbf{w}_2^{\mathbf{proj}},\ \dots, \mathbf{w}_{\mathbf{D_{proj}}}^{\mathbf{proj}}] \in \mathbb{R}^{D_{PCA\_S} \times D_{proj}}$, where $\mathbf{w}_{\mathbf{d}}^{\mathbf{proj}}$ denotes the eigenvector corresponding to the d-th largest positive eigenvalue of S'_B, $d = 1, 2, \dots, D_{proj}$. Calculate the between-class scatter matrix $S''_B \in \mathbb{R}^{D_{proj} \times D_{proj}}$ and the within-class scatter matrix $S''_W \in \mathbb{R}^{D_{proj} \times D_{proj}}$ by

$$S''_B = W_{proj}^t S'_B W_{proj} \tag{15}$$

$$S''_W = W_{proj}^t W_{PCA_S}^t S_W W_{PCA_S} W_{proj} \tag{16}$$

Adopt the optimization problem using the inverse Fisher discriminant criterion, we have

$$W_{IFDA} = \arg\min_{W} \frac{|W^t W_{proj}^t W_{PCA_S}^t S_W W_{PCA_S} W_{proj} W|}{|W^t W_{proj}^t W_{PCA_S}^t S_B W_{PCA_S} W_{proj} W|}$$

$$= \arg\min_{W} \frac{|W^t W_{proj}^t S_W' W_{proj} W|}{|W^t W_{proj}^t S_B' W_{proj} W|} = \arg\min_{W} \frac{|W^t S_W'' W|}{|W^t S_B'' W|} \quad (17)$$

Obtain $W_{IFDA} = \left[\mathbf{w_1}, \mathbf{w_2}, \ \ldots, \mathbf{w_{D_{proj}}}\right] \in \mathbb{R}^{D_{proj} \times D_{proj}}$ by solving the generalized eigenvalue problem

$$S_W'' \mathbf{w_d} = \lambda_d S_B'' \mathbf{w_d}, \ d = 1, 2, \ldots, D_{proj} \quad (18)$$

where $\boldsymbol{w}_d$ is the generalized eigenvector corresponding to the d-th smallest generalized eigenvalue λ_d.

Compute the optimal projection matrix $W_{OPT} \in \mathbb{R}^{RC \times D_{proj}}$ and a reduced vector $\mathbf{y_n} \in \mathbb{R}^{D_{proj}}$ as follows

$$W_{OPT}^t = W_{IFDA}^t W_{proj}^t W_{PCA\_S}^t \quad (19)$$

$$\mathbf{y}_n = W_{OPT}^t \mathbf{x}_n, \ n = 1, 2, \ldots, N. \quad (20)$$

3. 4. BDPCA plus LDA

Bi-Directional PCA (BDPCA) proposed by Zuo et al. (2006) is a generalization of two-dimensional principal component analysis (2DPCA) as shown by Yang et al. (2004), which directly computes eigenvectors of the scatter matrix without converting 2D matrices into 1D vectors. Rather than working in the row direction of an image like 2DPCA, BDPCA can reflect information between both rows and columns of an image. The BDPCA plus LDA performs LDA in the low-dimensional BDPCA subspace. The detailed algorithm and further discussion is given as follows.

3. 4. 1. Algorithm

(S1) (BDPCA): Construct the row total scatter matrix $S_T^{row} \in \mathbb{R}^{C \times C}$ and the column total scatter matrix $S_T^{col} \in \mathbb{R}^{R \times R}$ by

$$S_T^{row} = \frac{1}{NR} \sum_{n=1}^{N} (X_n - \bar{X})^t (X_n - \bar{X}), \ S_T^{col} = \frac{1}{NC} \sum_{n=1}^{N} (X_n - \bar{X})(X_n - \bar{X})^t \quad (21)$$

Find the projection matrix $W_{row} = \left[\mathbf{w_1^{row}}, \mathbf{w_2^{row}}, \ldots \mathbf{w_{D_{row}}^{row}}\right]$ and $W_{col} = [\mathbf{w_1^{col}}, \mathbf{w_2^{col}}, \ldots \mathbf{w_{D_{col}}^{col}}]$ according to the largest eigenvalues of S_T^{row} and S_T^{col}, respectively.

(S2) (BDPCA feature extraction):
Obtain BDPCA feature matrix $Y_n^{BDPCA} \in \mathbb{R}^{D_{col} \times D_{row}}$ of image matrix X_n $(n = 1, 2, \ldots, N)$ by

$$Y_n^{BDPCA} = W_{col}^t X_n W_{row} \quad (22)$$

Let $\mathbf{y_n^{BDPCA}} \in \mathbb{R}^{D_{col} D_{row}}$ be a column-vector representation of Y_n^{BDPCA} such that $\mathbf{y_n^{BDPCA}}(r * D_{row} + c) = Y_n^{BDPCA}(r, c)$, $n = 1, 2, \ldots, N$, $r = 0, 1, \ldots, D_{row} - 1$, $c = 0, 1, \ldots, D_{col} - 1$. Compute the mean image of class k: $\mathbf{\bar{y}_k^{BDPCA}} = \frac{1}{N_k} \sum_{\mathbf{y^{BDPCA}} \in class\ k} \mathbf{y^{BDPCA}}$, $k = 1, 2, \ \ldots, K$, and the mean image of all samples: $\mathbf{\bar{y}^{BDPCA}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y_n^{BDPCA}}$.

(S3) (LDA): Constructthe between-class scatter matrix $S_B^{BDPCA} \in \mathbb{R}^{D_{col}D_{row} \times D_{col}D_{row}}$ and the within-class scatter matrix$S_W^{BDPCA} \in \mathbb{R}^{D_{col}D_{row} \times D_{col}D_{row}}$in the BDPCA subspace by

$$S_B^{BDPCA} = \sum_{k=1}^{K} N_k (\bar{\mathbf{y}}_{\mathbf{k}}^{\mathbf{BDPCA}} - \bar{\mathbf{y}}^{\mathbf{BDPCA}})(\bar{\mathbf{y}}_{\mathbf{k}}^{\mathbf{BDPCA}} - \bar{\mathbf{y}}^{\mathbf{BDPCA}})^t \quad (23)$$

$$S_W^{BDPCA} = \sum_{k=1}^{K} \sum_{\mathbf{y}^{\mathbf{BDPCA}} \in class\ k} (\mathbf{y}^{\mathbf{BDPCA}} - \bar{\mathbf{y}}_{\mathbf{k}}^{\mathbf{BDPCA}})(\mathbf{y}^{\mathbf{BDPCA}} - \bar{\mathbf{y}}_{\mathbf{k}}^{\mathbf{BDPCA}})^t \quad (24)$$

Let $D_{LDA} = rank(S_B^{BDPCA})$, find the projection matrix $W_{OPT} = [\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_{D_{LDA}}}] \in \mathbb{R}^{D_{col}D_{row} \times D_{LDA}}$ by solving the generalized eigenvalue problem

$$S_B^{BDPCA} \mathbf{w}_d = \lambda_d S_W^{BDPCA} \mathbf{w}_d, \ d = 1, 2, \ldots, D_{LDA} \quad (25)$$

(S4) (Feature representation):
The final feature vector $\mathbf{y}_n \in \mathbb{R}^{D_{LDA}}$ is defined by

$$\mathbf{y}_n = W_{OPT}^t \mathbf{y}_n^{\mathbf{BDPCA}}, \ n = 1, 2, \ldots, N. \quad (26)$$

4. Experiments

The experiments were carried out using three face databases: the JAFFE database (Web-1) as shown by Lyons et al. (1998), the ORL database (Web-2) as shown by Samaria et al. (1994) and the FEI database (Web-3). The specifications for each of the face databases are further described below.

4. 1. 1. The JAFFE Database

The JAFFE database is a Japanese face data base which consists of 213 face images from 10 females. There are seven different facial expressions. Each individual has two to four images for each expression. All individuals are in an upright, front position. The original size of each image is 256×256 with 256 possible gray levels for each pixel. In our experiments, we cropped and resized them to 112×92 pixels. Only one image per facial expression was selected for one individual, so there are 7 images per individual and 70 images in total. The seven images of one individual are shown in Fig. 1.

Fig.1. Sample images from the JAFFE database

4. 1. 2. The ORL Database

The ORL database consists of 400 face images from 40 individuals, including 36 males and 4 females. Each contributes 10 different images. Some images were taken at different sessions, and there are variations in facial expressions and facial details. All the images were taken against a dark homogenous background with the individuals in an upright, front position (with a tolerance for tilting and rotation of the face up to 20 degrees).The size of each image is 112×92 with 256 possible gray levels for each pixel, we resized them to 56×46 pixels. The ten images of one individual from the ORL database are shown in Fig. 2.

Fig. 2.Sample images from the ORL database

4. 1. 3. The FEI Database

The FEI face database is a Brazilian face database that contains 2800 face images from 200 individuals, including 100 males and 100 females. Each contributes 14 images. All images are colorful and taken against a white homogenous background in an upright frontal position with an individual having rotation of up to 180 degrees. Scale might vary about 10% and the original size of each image is 640x480. In our experiments, we cropped and resized them to 56×46 pixels, and converted them to 256 gray levels. Only first 11 images in different rotation degrees were selected from each individual, so there are 2200 images in total. The eleven images of one individual used in the experiments from the FEI are shown in Fig. 3.

Fig. 3. Sample images from the FEI database

The information of images used in the experiments from three face databases are summarized in Table 1.

Table 1. A summary of the face databases

Name	Image Size (pixels)	Number				Variations			
		Total images	Images/individual	Individuals	Male/ Female	Expression	Facial details	Illumination	Pose
JAFFE	112×92	70	7	10	0 / 7	Yes	No	No	No
ORL	56×46	400	10	40	36 / 4	Yes	Yes	No	Yes
FEI	56×46	2200	11	200	100 / 100	No	No	No	Yes

4. 2. Performance Evaluation

We evaluated the performance of the subspace LDA methods on the aforementioned databases. The Euclidean distance and the nearest neighbour classifier is adopted. The recognition rate is calculated as the ratio of number of successful recognition and the total number of test images. We tested under different number of images per individual by splitting all images of each class into two subsets, such that there are k imagesper individual in training sets and the remaining $\frac{N}{K} - k$ images per individual are in the testing sets, $k = 1, \ldots, \frac{N}{K} - 1$. We tried all possible splits and report the average recognition rate for each database. All the experiments were carried out on an Intel Core i3-2100 computer with an 8GB RAM and tested on a Matlab platform (version: R2012b). The results are given in Tables 2 to 4.

Table 2. Recognition rates (%) of the subspace LDA methods on the JAFFE database

Subspace LDA methods	**Size of the images**	**Number of training images per individual**					
		1	**2**	**3**	**4**	**5**	**6**
a. Fisherface	112 x 92	N/A	88.6	96.0	99.5	100	100
b. Complete PCA plus LDA	112 x 92	82.1	91.7	95.8	99.3	100	100
c. IDAface	112 x 92	82.1	88.8	90.0	96.4	98.8	98.8
d. BDPCA plus LDA	112 x 92	78.6	81.8	91.6	98.9	100	100

Table 3. Recognition rates (%) of LDA and the subspace LDA methods on the ORL database

Methods	Number of training images per individual								
	1	2	3	4	5	6	7	8	9
Fisherface	N/A	78.2	87.5	90.7	92.6	93.5	98.9	99.0	100
Complete PCA plus LDA	73.6	85.8	92.0	94.7	96.4	97.4	99.6	99.8	100
IDAface	73.6	78.5	87.1	91.2	93.5	94.8	99.3	99.6	100
BDPCA plus LDA	76.4	86.3	92.2	95.3	96.8	97.9	99.7	99.9	100
LDA	N/A	55.9	57.1	60.4	66.5	74.4	97.0	99.0	100

Table 4. Recognition rates (%) of LDA and the subspace LDA methods on the FEI database

Subspace LDA methods	Number of training images per individual									
	1	2	3	4	5	6	7	8	9	10
Fisherface	N/A	41.8	51.4	55.9	58.1	59.4	59.8	91.1	97.6	96.4
Complete PCA plus LDA	44.2	68.0	78.6	82.6	84.4	85.0	84.7	96.6	99.2	98.8
IDAface	44.2	46.0	65.5	67.0	67.9	68.1	67.9	92.9	98.0	98.0
BDPCA plus LDA	54.8	75.8	85.9	90.9	93.8	95.5	96.5	99.4	99.9	100
LDA	N/A	39.0	45.7	48.3	50.3	51.4	52.0	88.1	94.7	95.8

4. Conclusion

We have discussed the LDA and four subspace LDA methods for solving the small sample size problem encountered in face recognition. Experimental results, by testing on three well-known face image databases: JAFFE(Web-1), ORL(Web-2), and FEI(Web-3), show that the LDA method directly converts a 2d image into a larger column vector perform the worst. On the other hand, the BDPCA plus LDA method applies both row and column projection to reduce the size of matrix achieves the overall best results, whereas, the size selection of subspace projection by either PCA or LDA merits further studies.

Acknowledgements

This work is supported by Taiwan National Science Council Grant NSC 101-2221-E-007-125-MY3.

References

Belhumeur P.N., Hespanha J.P., Kriegman, D. (1997).“Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,”*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720.

Chen C.C., Chu H.T., Shieh Y.S. (2008). Face Image Retrieval By Projection-Based Features, “The 3rd Int. Conf. on Image Media Quality and Its Applications,” Kyoto, Japan, Sep.5-6, pp.138-144.

Chen L.F., Liao H.Y.M., Ko M.T., Lin J, C., Yu G.J. (2000).“A New LDA-Based Face Recognition System which can Solve the Small Sample Size Problem,” *Pattern Recognition*, vol. 33, no. 10, pp.1713-1726.

Fisher R.A. (1936). “The Use of Multiple Measurements in Taxonomic Problems,” *Annals of Eugenics*, vol. 7, no. 2, pp. 179-188.

Huang C.T. (2014). A Study on the Subspace LDA Methods for Solving the Small Sample Size Problem in Face Recognition, M.S. Thesis, National Tsing Hua University, January 2014.

Kirby M., Sirovich L. (1990).“Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108.

Li S.Z., Jain A.K. (2011). Handbook of Face Recognition, Springer.

Lyons M., Akamatsu S., Kamachi M., Gyoba J. (1998). "Coding Facial Expressions with Gabor Wavelets," *The third IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 200-205.

Martínez A.M., Kak A.C. (2001). "PCA versus LDA," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233.

Samal A., Iyengar P.A. (1992). "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65-77.

Samaria F.S., Harter A.C. (1994). "Parameterisation of a Stochastic Model for Human Face Identification," *The Second IEEE Workshop on Applications of Computer Vision*, pp. 138-142.

Sirovich L., Kirby M. (1987). "Low-Dimensional Procedure for the Characterization of Human Faces," *The Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519-524.

Turk M., Pentland A. (1991). "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86.

Yang J., Yang J.Y. (2003). "Why Can LDA Be Performed in PCA Transformed Space?" *Pattern Recognition*, vol. 36, no. 2, pp. 563-566.

Yang J., Zhang D., Frangi A.F., Yang J.Y. (2004). "Two-dimensional PCA: a New Approach to Appearance-based Face Representation and Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131-137.

Yu H., Yang J. (2001). "A Direct LDA Algorithm for High-dimensional Data - with Application to Face Recognition," *Pattern Recognition*, vol. 34, pp. 2067-2070.

Zhao W., Chellappa R., Phillips P.J., Rosenfeld A. (2003). "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458.

Zhuang X.S., Dai D.Q. (2007). "Improved Discriminant Analysis for High-dimensional Data and Its Application to Face Recognition," *Pattern Recognition*, vol. 40, no. 5, pp. 1570-1578.

Zuo W., Zhang D., Yang, J., Wang K. (2006). "BDPCA Plus LDA: A Novel Fast Feature Extraction Technique for Face Recognition," IEEE Trans. on Systems, Man, and Cybernetics, Part B: *Cybernetics*, vol. 36, no. 4, pp. 946-953.

Web sites:

Web-1: http://www.kasrl.org/jaffe.html consulted 1 March, 2015.

Web-2: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html, consulted 1 March, 2015.

Web-3: http://fei.edu.br/~cet/facedatabase.html, consulted 1 March, 2015.