2

Subspace Methods for Face Recognition: Singularity, Regularization, and Robustness

Wangmeng Zuo, Kuanquan Wang and Hongzhi Zhang
Harbin Institute of Technology
China

1. Introduction

Face recognition has been an important issue in computer vision and pattern recognition over the last several decades (Zhao et al., 2003). While human can recognize faces easily, automated face recognition remains a great challenge in computer-based automated recognition research. One difficulty in face recognition is how to handle the variations in expression, pose and illumination when only a limited number of training samples are available.

Currently, face recognition methods can be grouped into three categories, feature-based, holistic-based, and hybrid approaches (Zhao et al., 2003). Feature-based approaches, which extract local features such as the locations and local statistics of the eyes, nose, and mouth, had been investigated in the beginning of the face recognition research (Kanade, 1973). Recently, with the introduction of elastic bunch graph matching (Wiskott, 1997) and local binary pattern (Timo, 2004), local feature-based approaches have shown promising results in face recognition. Holistic-based approaches extract a holistic representation of the whole face region, and have robust recognition performance under noise, blurring, and partial occlusion. After the introduction of Eigenfaces (Turk & Pentland, 1991) and Fisherfaces (Belhumeur et al., 1997), holistic-based approaches were extensively studied and widely applied to face recognition. Motivated by human perception system, hybrid approaches use both local feature and the whole face region for face recognition, and thus are expected to be potentially effective in improving recognition accuracy.

In holistic-based face recognition, feature extraction is fundamental, which can be revealed from three aspects. First, the input facial image is high dimensional and most current recognition approaches suffer from the "curse of dimensionality" problem. Thus a feature extraction step is necessary. Second, facial image usually contains less discriminative or unfavorable information for recognition (e.g., illumination). By making use of feature extraction, this information can be efficiently suppressed while retaining discriminative information. Third, feature extraction can greatly reduce the dimensionality of facial image, and this reduces the system's memory and computational requirements.

Subspace method, which aims to reduce the dimension of the data while retaining the statistical separation property between distinct classes, has been a natural choice for facial feature extraction. Face images, however, are generally high dimensional and their within-class variations is much larger than the between-class variations, which will cause the serious performance degradation of classical subspace methods. By far, various subspace methods have been proposed and applied to face recognition.

Source: State of the Art in Face Recognition, Book edited by: Dr. Mario I. Chacon M.,
ISBN -3-902613-42-4, pp. 250, January 2009, I-Tech, Vienna, Austria

1.1 Previous work

At the beginning, linear unsupervised method, such as principal component analysis (PCA), was used to extract the holistic feature vectors for facial image representation and recognition (Turk & Pentland, 1991). Other unsupervised methods, such as independent component analysis (ICA) and non-negative matrix factorization (NMF), have been subsequently applied to face recognition (Bartlett et al., 2002; Zafeiriou et al., 2006).

Since the unsupervised methods do not utilize the class label information in the training stage, it is generally believed that the supervised methods are more effective in dealing with recognition problems. Fisher linear discriminant analysis (LDA), which aims to find a set of optimal discriminant vectors that map the original data into a low-dimensional feature space, is then gaining popularity in face recognition. In 1996, Fisher linear discriminant analysis was applied to face recognition, and subsequently was developed into one of the most famous face recognition approaches, Fisherfaces (Swets & Weng, 1996; Belhumeur et al., 1997). In face recognition, the data dimensionality is much higher than the size of the training set, leading to the small sample size problem (the SSS problem). Currently there are two popular strategies to solve the SSS problem, the transform-based and the algorithm-based. The transform-based strategy first reduces the dimensions of the original image data and then uses LDA for feature extraction, while the algorithm-based strategy finds an algorithm to circumvent the SSS problem (Yang & Yang, 2003; Yu & Yang, 2001).

Face recognition usually is highly complex and can not be regarded as a linear problem. In the last few years, a class of nonlinear discriminant analysis techniques named as kernel discriminant analysis has been widely investigated for face recognition. A number of kernel-methods, such as kernel principal component analysis (KPCA), kernel Fisher's discriminant analysis, complete kernel Fisher discriminant (CKFD), and kernel direct discriminant analysis (KDDA), have been developed (Liu, 2004; Yang, 2002; Yang et al., 2005b; Lu et al., 2003). Most recently, manifold learning methods, such as isometric feature mapping (ISOMAP), locally linear embedding (LLE), and Laplacian eigenmaps, have also shown great potential in face recognition (Tenenbaum et al., 2000; Roweis & Saul, 2000; He et al., 2005).

As a generalization of vector-based methods, a number of tensor discrimination technologies have been proposed. The beginning of tensor discrimination technology can be traced back to 1993, where a 2D image matrix based algebraic feature extraction method is proposed for image recognition (Liu et al., 1993). As a new development of the 2D image matrix based straightforward projection technique, a two-Dimensional PCA (2DPCA) approach was suggested for face representation and recognition (Yang et al., 2004). To further reduce computational cost, researchers had developed several BDPCA and generalized low rank approximations of matrices (GLRAM) approaches (Ye, 2004; Zuo et al., 2005a). Motivated by multilinear generalization of singular vector decomposition (Lathauwer et al., 2000), a number of alterative supervised and unsupervised tensor analysis methods have been proposed for facial image or image sequence feature extraction (Tao et al., 2005; Yan et al., 2007).

1.2 Organization of this chapter

Generally, there are three issues which should be addressed in the development of subspace methods for face recognition, singularity, regularization, and robustness. First, the dimensionality of facial image usually is higher than the size of the available training set,

which results in the singularity of the scatter matrices and causes the performance degradation (known as the SSS problem). So far, considerable research interests have been given to solve the SSS problem. Second, another unfavorable effect of the SSS problem is that, a limited sample size can cause poor estimation of the scatter matrices, resulting in an increase in the classification error. Third, noisy or partially occluded facial image may be inevitable during the capture and communication stage, and thus the robust recognition should be addressed in the development of subspace methods.

In this chapter, we introduce the recent development of subspace-based face recognition methods in addressing these three problems. First, to address the singularity problem, this chapter proposes a fast feature extraction technique, Bi-Directional PCA plus LDA (BDPCA+LDA), which performs LDA in the BDPCA subspace. Compared with the PCA+LDA framework, BDPCA+LDA needs less computational and memory requirements, and can achieve competitive recognition accuracy. Second, to alleviate the over-fitting to the training set, this chapter suggests a post-processing approach on discriminant vectors, and theoretically demonstrates its relationship with the image Euclidean distance method (IMED). Third, to improve the robustness of subspace method over noise and partial occlusion, this chapter presents an iteratively reweighted fitting of the Eigenfaces method (IRF-Eigenfaces), which first defines a generalized objective function and then uses the iteratively reweighted least-squares (IRLS) fitting algorithm to extract the feature vector by minimizing the generalized objective function. Finally, two popular face databases, the AR and the FERET face databases, are used to evaluate the performance the proposed subspace methods.

2. BDPCA+LDA: a novel method to address the singular problem

In face recognition, classical LDA always encounters the SSS problem, where the data dimensionality is much higher than the size of the training set, leading to the singularity of the within-class scatter matrix $\mathbf{S}_w$. A number of approaches have been proposed to address the SSS problem. One of the most successful approaches is subspace LDA which uses a dimensionality reduction technique to map the original data to a low-dimensional subspace. Researchers have applied PCA, latent semantic indexing (LSI), and partial least squares (PLS) as pre-processors for dimensionality reduction (Belhumeur et al., 1997; Torkkola, 2001; Baeka & Kimb, 2004). Among all the subspace LDA methods, over the past decade, the PCA plus LDA approach (PCA+LDA), where PCA is first applied to eliminate the singularity of $\mathbf{S}_w$, and then LDA is performed in the PCA subspace, has received significant attention (Belhumeur et al., 1997). The discarded null space of $\mathbf{S}_w$, however, may contain some important discriminant information and cause the performance deterioration of Fisherfaces. Rather than discarding the null space of $\mathbf{S}_w$, Yang proposed a complete PCA+LDA method which simultaneously considered the discriminant information both in the range space and the null space of $\mathbf{S}_w$ (Yang & Yang, 2003).

In this section, we introduce a fast subspace LDA technique, Bi-Directional PCA plus LDA (BDPCA+LDA). BDPCA, which assumes that the transform kernel of PCA is separable, is a natural extension of classical PCA and a generalization of 2DPCA (Yang et al., 2004). The separation of the PCA kernel has at least three main advantages: lower memory requirement, faster training and feature extraction speed.

2.1 Linear discriminant analysis

Let $\mathbf{M}$ be a set of data, $\mathbf{M} = \{\mathbf{x}_{11}, \cdots, \mathbf{x}_{1n_1}, \cdots, \mathbf{x}_{ij}, \cdots, \mathbf{x}_{Cn_C}\}$, where $\mathbf{x}_{ij}$ is the jth training sample of the ith class, and n_i is the number of samples of the ith class, C is the number of classes. The sample $\mathbf{x}_{ij}$ is a one-dimensional vector or a vector representation of the corresponding image $\mathbf{X}_{ij}$. LDA and PCA are two classical dimensionality reduction techniques. PCA, an optimal representation method in a minimization of mean-square error sense, has been widely used for the representation of shape, appearance, and video (Jolliffe, 2001). LDA is a linear dimensionality reduction technique which aims to find a set of the optimal discriminant vectors by maximizing the class separability criterion (Fukunaga, 1990). In the field of face recognition, LDA is usually assumed more effective than PCA because LDA aims to find the optimal discriminant directions.

Two main tasks in LDA are calculation of the scatter matrices, and selection of the class separability criterion. Most LDA algorithms involve the simultaneous maximization of the trace of a scatter matrix and minimization of the trace of another matrix. LDA usually makes use of two scatter matrices, such as the within-class scatter matrix $\mathbf{S}_w$ and the between-class scatter matrix $\mathbf{S}_b$. The within-class scatter matrix $\mathbf{S}_w$, the scatter of samples around their class mean vectors, is defined as

$$\mathbf{S}_w = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)^T \ . \tag{1}$$

The between-class scatter matrix $\mathbf{S}_b$, the scatter of class mean vectors around the global mean vector, is defined as

$$\mathbf{S}_b = \frac{1}{N}\sum_{i=1}^{C} n_i(\overline{\mathbf{x}}_i - \overline{\mathbf{x}})(\overline{\mathbf{x}}_i - \overline{\mathbf{x}})^T \ , \tag{2}$$

where $\overline{\mathbf{x}} = \sum_{i=1}^{C}\sum_{j=1}^{n_i}\mathbf{x}_{ij}/N$ is the global mean vector, $\overline{\mathbf{x}}_i = \sum_{j=1}^{n_i}\mathbf{x}_{ij}/n_i$ is the mean vector of class i, and $N = \sum_{i=1}^{C} n_i$ is the total number of training samples.

The most famous class separability criterion is the Fisher's discriminant criterion

$$J_F(\mathbf{w}) = \frac{\mathbf{w}\mathbf{S}_b\mathbf{w}}{\mathbf{w}\mathbf{S}_w\mathbf{w}} \ . \tag{3}$$

The set of discriminant vectors $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_{d_{LDA}}]$ corresponding to the maximization of the Fisher's discriminant criterion can be obtained by solving the generalized eigenvalue problem $\mathbf{S}_b\mathbf{W} = \mathbf{S}_w\mathbf{W}\mathbf{\Lambda}$. Since both $\mathbf{S}_b$ and $\mathbf{S}_w$ are symmetric matrices, the *simultaneous diagonalization* technique can be used to calculate the set of discriminant vectors $\mathbf{W}$.

2.1.1 Simultaneous diagonalization

Fig. 1 uses a three-class problem to illustrate the procedure of simultaneous diagonalization in computing the discriminant vectors of LDA. The distribution of each class and the distributions of within- and between-class scatter are depicted in Fig. 1(a) and (b). Simultaneous diagonalization tries to find a transformation matrix $\mathbf{\Phi}$ that satisfies $\mathbf{\Phi}^T\mathbf{S}_w\mathbf{\Phi}=\mathbf{I}$ and $\mathbf{\Phi}^T\mathbf{S}_b\mathbf{\Phi}=\mathbf{\Lambda}_g$, where $\mathbf{I}$ is an identity matrix and $\mathbf{\Lambda}_g$ is a diagonal matrix.

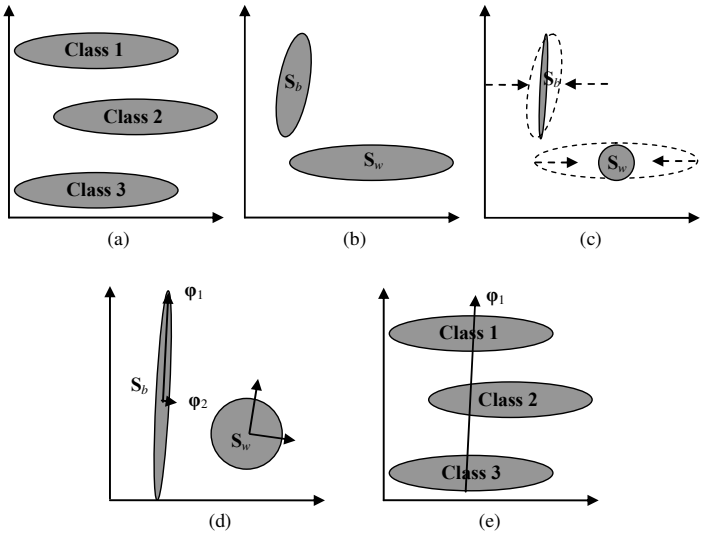


Fig. 1. Procedure of simultaneous diagonalization: (a) The distributions of the three-class problem, (b) The within-and between-class distributions, (c) Whitening the within-class distribution, and correspondingly transform the between-class distribution. (d) Calculating the eigenvectors of the transformed between-class distribution. (e) Illustration of the first discriminant vector.

The procedure of simultaneous diagonalization contains three steps:

Step 1. Whitening $\mathbf{S}_w$. PCA is used to whiten the within-class distribution to an isotropic distribution by a transformation matrix $\mathbf{\Theta}_{wh}$. Then, matrix $\mathbf{\Theta}_{wh}$ is used to transform the between-class scatter $\hat{\mathbf{S}}_b=\mathbf{\Theta}_{wh}{}^T\mathbf{S}_b\mathbf{\Theta}_{wh}$.

Step 2. Calculation of the eigenvectors $\mathbf{\Psi}$ and eigenvalues $\mathbf{\Lambda}_g$ of $\hat{\mathbf{S}}_b$.

Step 3. Computation of the transformation matrix $\mathbf{\Phi}=\mathbf{\Theta}_{wh}\mathbf{\Psi}$, where $\mathbf{\Phi}=[\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2]$ is the set of generalized eigenvectors of $\mathbf{S}_w$ and $\mathbf{S}_b$.

2.2 BDPCA+LDA: algorithm

2.2.1 Bi-directional PCA

To simplify our discussion, in the following, we adopt two representations of an image, $\mathbf{X}$ and $\mathbf{x}$, where $\mathbf{X}$ is a representation of an image matrix and $\mathbf{x}$ is a representation of an image vector. $\mathbf{X}$ and $\mathbf{x}$ represent the same image.

Given a transform kernel (e.g., principal component vector) $\mathbf{w}_i$, an image vector $\mathbf{x}$ can be projected into $\mathbf{w}_i$ by $y_i = \mathbf{w}_i^T\mathbf{x}$. In image transform, if the transform kernel is *product-separable*, the image matrix $\mathbf{X}$ can be projected into $\mathbf{w}_i$ equivalently by $y_i = \mathbf{w}_{i,C}^T\mathbf{X}\mathbf{w}_{i,R}$, where $\mathbf{w}_{i,C}$ and $\mathbf{w}_{i,R}$ are the corresponding column transform kernel and row transform kernel of $\mathbf{w}_i$. In PCA, assuming all the eigenvectors $\mathbf{W}=[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d]$ are *product-separable*, there are two equivalent ways to extract the feature of an image $\mathbf{x}$, $\mathbf{y} = \mathbf{w}^T\mathbf{x}$ (vector-based way) and $\mathbf{Y} = \mathbf{W}_C^T\mathbf{X}\mathbf{W}_R$ (matrix-based way), where $\mathbf{W}_C$ and $\mathbf{W}_R$ are the column and row projection matrices.

2DPCA assumes the column projection matrix $\mathbf{W}_C$ is an $m \times m$ identity matrix, and the criterion of classical PCA will degenerate to

$$J(\mathbf{w}) = \mathbf{w}^T \mathbf{G}_t \mathbf{w}\,, \tag{4}$$

where $\mathbf{w}$ is a unitary column vector, $\mathbf{w}^T\mathbf{w} = 1$, and $\mathbf{G}_t$ is the image covariance matrix defined as $\mathbf{G}_t = \sum_{i=1}^{N} (\mathbf{X}_i - \bar{\mathbf{X}})^T (\mathbf{X}_i - \bar{\mathbf{X}})$. Compared with PCA, 2DPCA has several significant advantages. First, 2DPCA is simpler and more straightforward to use for image feature extraction. Second, experimental results consistently show that 2DPCA is better than PCA in terms of recognition accuracy. Third, 2DPCA is computationally more efficient than PCA and significantly improve the speed of image feature extraction (Yang et al., 2004).

Bi-Directional PCA (BDPCA) extracts representative feature from image $\mathbf{X}$ by $\mathbf{Y} = \mathbf{W}_C^T \mathbf{X} \mathbf{W}_R$ yet it is difficult to simultaneously determine optimal $\mathbf{W}_C$ and $\mathbf{W}_R$ in an analytic framework. However, a number of alternative approaches have been proposed to compute the optimal column and row projection matrices $\mathbf{W}_C$ and $\mathbf{W}_R$. In the following, we summary the three main strategies for dealing with this:

1. The Hierarchical Strategy (Yang et al., 2005a). Hierarchical strategy adopts a two-step framework to calculate $\mathbf{W}_C$ and $\mathbf{W}_R$. First a 2DPCA is performed in horizontal direction and the second 2DPCA is performed on the row-compressed matrix in vertical direction (**H1**), as shown in Fig. 2(a). It is obvious that we can adopt an alternative method, first perform 2DPCA in vertical direction and then in horizontal direction (**H2**).
2. The Iterative Strategy. In (Ye, 2005), Ye proposed an iterative procedure for computing $\mathbf{W}_C$ and $\mathbf{W}_R$. After the initialization of $\mathbf{W}_{C0}$, the procedure repeatedly first updates $\mathbf{W}_R$ according to $\mathbf{W}_C$, and then updates $\mathbf{W}_C$ according to $\mathbf{W}_R$ until convergence (**I1**), as shown in Fig. 2(b). Theoretically, this procedure can only be guaranteed to be convergent to locally optimal solution of $\mathbf{W}_C$ and $\mathbf{W}_R$. Their experimental results also show that, for image data with some hidden structure, the iterative algorithm may converge to the global solution, but this assertion does not always hold.
3. The Independence Assumption (Zuo et al., 2006). One disadvantage of the hierarchical strategy is that are always confronted with the choice of **H1** or **H2**. Assuming that the computing of $\mathbf{W}_R$ and the computing of $\mathbf{W}_C$ are independent, $\mathbf{W}_C$ and $\mathbf{W}_R$ can be computed by solving two 2DPCA problems independently (**I2**), as shown in Fig. 2(c). Experimental results show that, in facial feature extraction, **H1**, **H2**, **I1** and **I2** have similar recognition performance, and **H1**, **H2**, and **I2** require less training time.

In the following, we use the third strategy to explain the procedure of BDPCA. Given a training set $\{\mathbf{X}_1, \cdots, \mathbf{X}_N\}$, N is the number of the training images, and the size of each image matrix is $m \times n$. By representing the ith image matrix $\mathbf{X}_i$ as an m-set of $1 \times n$ row vectors

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_i^1 \\ \mathbf{x}_i^2 \\ \vdots \\ \mathbf{x}_i^m \end{bmatrix}, \tag{5}$$

we adopt Yang's approach (Yang et al, 2004) to define the row total scatter matrix

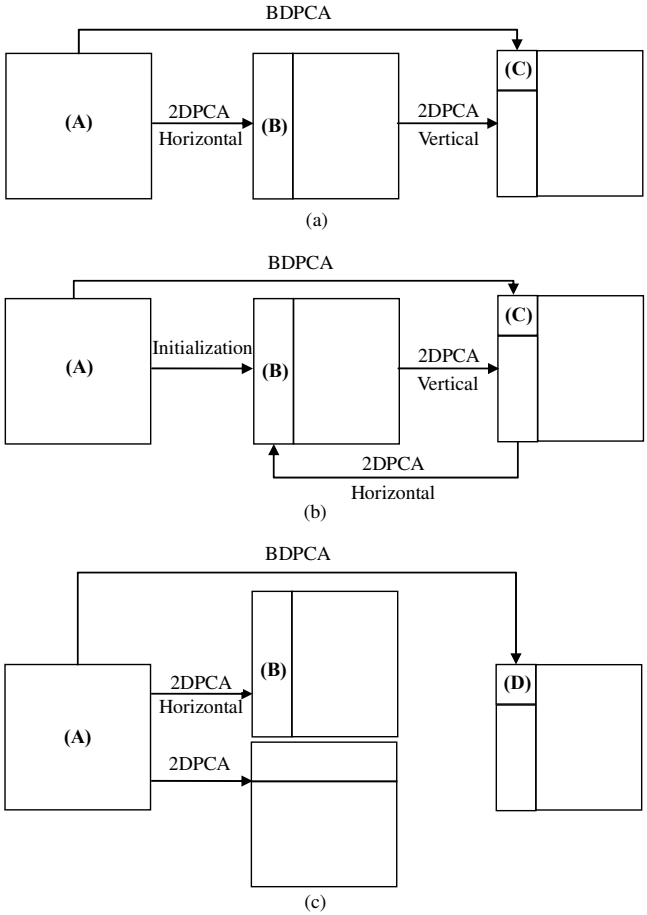


Fig. 2. Illustration of the three strategies in calculating the column and row transformation matrix

$$\mathbf{S}_t^{row} = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{j=1}^{m}(\mathbf{x}_i^j - \bar{\mathbf{x}}^j)^T(\mathbf{x}_i^j - \bar{\mathbf{x}}^j) = \frac{1}{Nm}\sum_{i=1}^{N}(\mathbf{X}_i - \bar{\mathbf{X}})^T(\mathbf{X}_i - \bar{\mathbf{X}}), \tag{6}$$

where $\mathbf{x}_i^j$ and $\bar{\mathbf{x}}_i^j$ denotes the jth row of sample $\mathbf{X}_i$ and mean matrix $\bar{\mathbf{X}}$, respectively. We choose the row eigenvectors corresponding to the first k_{row} largest eigenvalues of $\mathbf{S}_t^{row}$ to construct the row projection matrix $\mathbf{W}_r$

$$\mathbf{W}_r = [\mathbf{w}_1^{row}, \mathbf{w}_2^{row}, \cdots, \mathbf{w}_{k_{\text{row}}}^{row}], \tag{7}$$

where $\mathbf{w}_i^{row}$ denotes the row eigenvector corresponding to the ith largest eigenvalues of $\mathbf{S}_t^{row}$. Similarly, by treating an image matrix $\mathbf{X}_i$ as an n-set of m×1 column vectors

$$\mathbf{X}_i = [\mathbf{x}_i^1 \quad \mathbf{x}_i^2 \quad \cdots \quad \mathbf{x}_i^n], \tag{8}$$

we define the column total scatter matrix

$$\mathbf{S}_t^{col} = \frac{1}{Nn}\sum_{i=1}^{N}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T . \quad (9)$$

We then choose the column eigenvectors corresponding to the first k_{col} largest eigenvalues of $\mathbf{S}_t^{col}$ to construct the column projection matrix $\mathbf{W}_c$

$$\mathbf{W}_c = [\mathbf{w}_1^{col}, \mathbf{w}_2^{col}, \cdots, \mathbf{w}_{k_{\text{col}}}^{col}] , \quad (10)$$

where $\mathbf{w}_i^{col}$ is the column eigenvector corresponding to the ith largest eigenvalues of $\mathbf{S}_t^{col}$. Finally we use the transformation

$$\mathbf{Y} = \mathbf{W}_c{}^T \mathbf{X} \mathbf{W}_r , \quad (11)$$

to extract the feature matrix $\mathbf{Y}$ of image matrix $\mathbf{X}$.

2.2.2 BDPCA+LDA

BDPCA+LDA is an LDA approach that is applied on a low-dimensional BDPCA subspace, and thus can be used for fast facial feature extraction. Since less time is required to map an image matrix to BDPCA subspace, BDPCA+LDA is, at least, computationally faster than PCA+LDA.

BDPCA+LDA first uses BDPCA to obtain feature matrix $\mathbf{Y}$. The feature matrix $\mathbf{Y}$ is then transformed into feature vector $\mathbf{y}$ by concatenating the columns of $\mathbf{Y}$. The LDA projector $\mathbf{W}_{\text{LDA}} = [\varphi_1, \varphi_2, \cdots, \varphi_m]$ is calculated by maximizing Fisher's criterion:

$$J(\varphi) = \frac{\varphi^T \mathbf{S}_b \varphi}{\varphi^T \mathbf{S}_w \varphi} , \quad (12)$$

where φ_i is the generalized eigenvector of $\mathbf{S}_b$ and $\mathbf{S}_w$ corresponding to the ith largest eigenvalue λ_i

$$\mathbf{S}_b \varphi_i = \lambda_i \mathbf{S}_w \varphi_i , \quad (13)$$

and $\mathbf{S}_b$ is the between-class scatter matrix of $\mathbf{y}$

$$\mathbf{S}_b = \frac{1}{N}\sum_{i=1}^{C} N_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T , \quad (14)$$

and $\mathbf{S}_w$ is the within-class scatter matrix of $\mathbf{y}$,

$$\mathbf{S}_w = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i} (\mathbf{y}_{i,j} - \boldsymbol{\mu}_i)(\mathbf{y}_{i,j} - \boldsymbol{\mu}_i)^T , \quad (15)$$

where N_i, $\mathbf{y}_{i,j}$ and $\boldsymbol{\mu}_i$ are the number of feature vectors, the jth feature vector and the mean vector of class i, C is the number of classes, and $\boldsymbol{\mu}$ is the mean vector of all the feature vectors.

In summary, the main steps in BDPCA+LDA feature extraction are to first transform an image matrix $\mathbf{X}$ into BDPCA feature subspace $\mathbf{Y}$ by Eq. (11), and map $\mathbf{Y}$ into its 1D representation $\mathbf{y}$ and then to obtain the final feature vector $\mathbf{z}$ by

$$\mathbf{z} = \mathbf{W}_{\mathrm{LDA}}^{\mathrm{T}}\mathbf{y} \ . \tag{16}$$

2.2.3 Advantages over the existing PCA plus LDA framework

We compare the BDPCA+LDA and the PCA+LDA face recognition frameworks in terms of their computational and memory requirements. It is worth noting that the computational requirements are considered in two phases, training and testing.

Method	Memory Requirements		Computation Requirements	
	Projector	Feature prototypes	Training	Testing
PCA+LDA	$(m\times n)\times d_{\mathrm{LDA}}$ **Large**	$N\times d_{\mathrm{LDA}}$ **Same**	a) Calculating the projector: $O(N_p^3+d_{\mathrm{PCA}}^3)$ **Large** b) Projection: $N\times(m\times n)\times d_{\mathrm{LDA}}$ **Large**	c) Projection: $(m\times n)\times d_{\mathrm{LDA}}$ **Large** d) Distance calculation: $N\times d_{\mathrm{LDA}}$ **Same**
BDPCA+LDA	$m\times k_{\mathrm{row}}+$ $n\times k_{\mathrm{col}}+$ $k_{\mathrm{col}}\times k_{\mathrm{row}}\times d_{\mathrm{LDA}}$ **Small**	$N\times d_{\mathrm{LDA}}$ **Same**	a) Calculating the projector: $O(m^3+n^3+d_{\mathrm{BDPCA}}^3)$ **Small** b) Projection: $N\times[m\times n\times\min(k_{\mathrm{row}}, k_{\mathrm{col}})$ $+k_{\mathrm{col}}\times k_{\mathrm{row}}\times\max(m+d_{\mathrm{LDA}},$ $n+d_{\mathrm{LDA}})]$ **Small**	c) Projection: $m\times n\times\min(k_{\mathrm{row}}, k_{\mathrm{col}})+$ $k_{\mathrm{col}}\times k_{\mathrm{row}}\times[\max(m,n)+d_{\mathrm{LDA}}]$ **Small** d) Distance calculation: $N\times d_{\mathrm{LDA}}$ **Same**

Table 1. Comparisons of computational and memory requirements of BDPCA+LDA and PCA+LDA

We first compare the computational requirement using the number of multiplications as a measurement of computational complexity. The training phase involves two computational tasks: a) calculation of the projector, and b) projection of images into feature prototypes. To calculate the projector, the PCA+LDA method must solve an $N\times N$ eigenvalue problem and then a $d_{\mathrm{PCA}}\times d_{\mathrm{PCA}}$ generalized eigenvalue problem, where N is the size if the training set and d_{PCA} is the dimension of the PCA subspace. In contrast, BDPCA+LDA must solve an $m\times m$, an $n\times n$ eigenvalue problem and a $d_{\mathrm{BDPCA}}\times d_{\mathrm{BDPCA}}$ generalized eigenvalue problem, where d_{BDPCA} is the dimension of BDPCA subspace. Since the complexity of an $M\times M$ eigenvalue problem is $O(M^3)$, the complexity of the PCA+LDA projector-calculation operation is $O(N^3+d_{\mathrm{PCA}}^3)$ whereas that of BDPCA+LDA is $O(m^3+n^3+d_{\mathrm{BDPCA}}^3)$. Assuming that m, n, d_{PCA} and d_{BDPCA} are smaller than the number of training samples N, in calculating the projector, BDPCA+LDA requires less computation than PCA+LDA to calculate the projector.

To project images into feature prototypes, we assume that the feature dimension of BDPCA+LDA and PCA+LDA is the same, d_{LDA}. For PCA+LDA, the number of multiplications is thus $N_p\times(m\times n)\times d_{\mathrm{LDA}}$. For BDPCA+LDA, the number of multiplications is less than $N_p\times(m\times n\times\min(k_{\mathrm{row}},k_{\mathrm{col}})+(k_{\mathrm{col}}\times k_{\mathrm{row}})\times\max(m+d_{\mathrm{LDA}}, n+d_{\mathrm{LDA}}))$, where N_p is the number

of prototypes. In this paper, we use all the prototypes for training, thus $N_p=N$. Assuming that $\min(k_{row}, k_{col})$ is much less than d_{LDA}, in the projection process, BDPCA+LDA also requires less computation than PCA+LDA.

In the test phase, there are two computational tasks: c) the projection of images into the feature vector, and d) the calculation of the distance between the feature vector and feature prototypes. In the following we compare the computational requirement of BDPCA+LDA and PCA+LDA in carrying out these two tasks. When projecting images into feature vectors, BDPCA+LDA requires less computation than PCA+LDA. Because the feature dimension of BDPCA+LDA and PCA+LDA is the same, in the similarity measure process, the computational complexity of BDPCA+LDA and PCA+LDA are equal. Taking these two tasks into account, BDPCA+LDA is also less computationally expensive than PCA+LDA in the testing phase.

The memory requirements of the PCA+LDA and BDPCA+LDA frameworks mainly depend on the size of the projector and the total size of the feature prototypes. The size of the projector of PCA+LDA is $d_{LDA}\times m\times n$. This is because the PCA+LDA projector contains d_{LDA} Fisherfaces, each of which is the same size as the original image. The BDPCA+LDA projector is in three parts, $\mathbf{W}_c$, $\mathbf{W}_r$ and $\mathbf{W}_{LDA}$. The total size of the BDPCA+LDA projector is $(k_{col}\times m)+(k_{row}\times n)+(d_{LDA}\times k_{col}\times k_{row})$, which is generally much smaller than that of PCA+LDA. Finally, because these two methods have the same feature dimensions, BDPCA+LDA and PCA+LDA have equivalent feature prototype memory requirements.

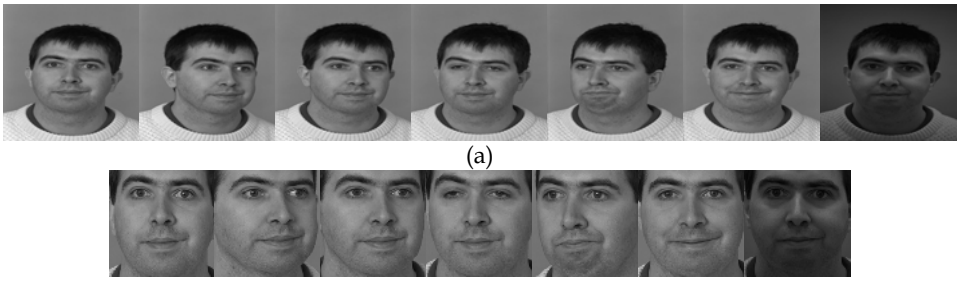
We have compared the computational and memory requirements of the BDPCA+LDA and PCA+LDA frameworks, as listed in Table 1. Generally, the BDPCA+LDA framework is superior to the PCA+LDA in both the computational and memory requirements.

2.3 BDPCA+LDA: experimental results

To evaluate the efficacy of BDPCA+LDA we make use of the FERET face database. The FERET face database is a US Department of Defense-sponsored face database and is one of the standard databases used in testing and evaluating face recognition algorithms (Phillips, 1998; Phillips et al., 2000). For our experiments, we chose a subset of the FERET database. This subset includes 1,400 images of 200 individuals (each individual contributing seven images). The seven images of each individual consist of three front images with varied facial expressions and illuminations, and four profile images ranging from ±15° to ±25° pose. The facial portion of each original image was cropped to a size of 80×80 and pre-processed using histogram equalization. Fig. 3 illustrates the seven images of one person and their corresponding cropped images.

We also compare BDPCA+LDA with other LDA-based methods, including Fisherfaces, Enhanced Fisher discriminant Model (EFM) (Liu & Wechsler, 1998), Discriminant Common Vectors (DCV) (Cevikalp et al., 2005), and D-LDA. The experimental setup is as follows. Since our aim is to evaluate the efficacy of feature extraction methods, we use a simple classifier, the nearest neighbor classifier. To reduce the variation of recognition results, we adopt the mean of 10 runs as the average recognition rate (ARR). All the experiments are carried out on an AMD 2500+ computer with 512Mb RAM and tested on the MATLAB platform (Version 6.5).

In our experiments, three images of each person are randomly chosen for training, while the remaining four images are used for testing. Thus, we obtain a training set of 600 images and a testing set of 800 images. In this way, we run the face recognition method 10 times and calculate the average recognition rate.



(a)

(b)

Fig. 3. Images of an individual in the FERET subset: (a) the original images, and (b) the cropped images.

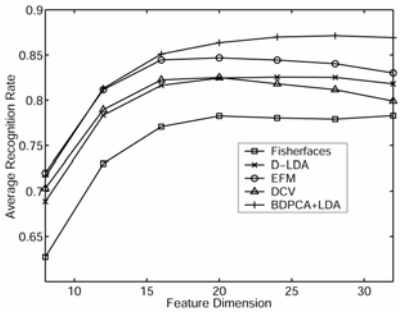


Fig. 4. Comparisons of the recognition rates obtained using different methods on the FERET subset.

Methods	Fisherfaces	D-LDA	EFM	DCV	BDPCA+LDA
Parameters	$[d_{PCA}, d]$	$[d_b, d]$	$[d_{PCA}, d_w, d]$	$[d]$	$[k_{col}, k_{row}, d]$
Values	[200, 20]	[60, 24]	[100, 100, 24]	[20]	[15, 5, 28]
ARR (%)	78.26%	82.56%	84.69%	82.51%	87.14%

Table 2. Recognition performance of five face recognition methods on the FERET database

Method	Time for Training (s)	Time for Testing (s)
PCA+LDA	254.2	36.2
BDPCA+LDA	57.5	26.3

Table 3. The total CPU time (s) for training and testing on the FERET database

We compare the recognition rates obtained using BDPCA+LDA, Fisherfaces, EFM, DCV and D-LDA, as shown in Fig. 4. We also list the optimal parameter values of each method and its maximum ARR in Table 2. The maximum ARR of BDPCA+LDA is 87.14%, higher than the ARRs of the other four methods.

Table 3 shows the total CPU time of PCA+LDA (EFM) and BDPCA+LDA in the training phase and the testing phase. BDPCA+LDA is much faster than EFM in both the training and testing phases.

We compare the computational and memory requirements of BDPCA+LDA and PCA+LDA (EFM). In Section 2.2.3, based on a number of assumptions, we assert that BDPCA+LDA is superior to PCA+LDA in the computational and memory requirements. We then check the

correctness of these assumptions. The size of the training set is 600, much higher than the size of row vector (80) or column vector (80). The feature dimension of EFM is 24, much higher than k_{row} (5). Thus all these assumptions are satisfied. Table 4 shows the computational and memory requirements of BDPCA+LDA and EFM. BDPCA+LDA needs less computational and memory requirements than EFM.

Method	Memory Requirements			Computation Requirements	
	Projector	Feature prototypes	Total	Training	Testing
PCA+LDA	(112×92)×39 =*401856*	200×39 =*7800*	*409656*	a) Calculating the projector: $O(200^3+160^3) \approx 12096000$ b) Projection: 200×(112×92)×39 =*80371200* Total=*92467200*	c) Projection: (112×92)×39=401856 d) Distance calculation: 200×39=*7800* Total=*409656*
BDPCA+LDA	112×12+92×4+(12×4)×39 =*3584*	200×39 =*7800*	*11384*	a) Calculating the projector: $O(112^3+92^3+(12\times4)^3) \approx 2294208$ b) Projection: 200×[112×92×4+12×4×(112+39)]=*9692800* Total=*11987008*	c) Projection: 112×92×4+4×12×(112+39) =*48464* d) Distance calculation: 200×39 =*7800* Total=*56264*

Table 4. Comparisons of computational and memory requirements of BDPCA+LDA and PCA+LDA on the FERET subset

It should be noted that, the training complexity of BDPCA+LDA is $O(N)$, whereas that of PCA+LDA is $O(N^3)$, where N is the size of training set. This property implies that, when the size of the training set is high, BDPCA+LDA would be more superior to PCA+LDA in terms of computational requirement.

3. Regularization of LDA: a post-processing approach

Despite the great success of LDA in face recognition, there still exist some potential issues deserving further investigation. One is that the discriminant vectors may be over-fitted to the training set, and are very noisy and wiggly in appearance. Another disadvantage of traditional LDA is that it does not take into account the spatial relationship of pixels. Since the inaccurate location and small perturbation is unavoidable in face detection and recognition, spatial information would be helpful to improve the robustness of the recognition performance. In addressing this issue, Wang et al. (2005) proposed an image Euclidean distance (IMED) method, where a 2D-Gaussian function is used to model the effect of neighbor pixels.

In the following, we first introduce a post-processed LDA-based method, and then demonstrate the equivalence of IMED and the post-processing approach. Finally, the FERET face database is used to evaluate the performance of post-processed LDA.

3.1 Post-processed LDA

Post-processing on discriminant vectors is effective in the improvement of the recognition performance of LDA-based face recognition methods. In this section, we first briefly summarize the post-processing approach, and then present an example of the post-processing approach, post-processed enhanced Fisher's model (PEFM).

3.1.1 Post-processing approach

A post-processing approach, 2D-Gaussian filtering, has been introduced to perform on the discriminant vectors (Wang et al., 2005; Zuo et al., 2005b). 2D-Gaussian filter is an ideal filter in the sense that it reduces the magnitude of high spatial frequency in an image and has been widely applied in image smoothing and denoising. In face recognition, where the discriminant vector can be mapped to a 2D image, Gaussian filtering is used to post-process the discriminant images and reduce noise. 2D-Gaussian function is defined as

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}, \tag{17}$$

where σ is the standard deviation. First a 2D-Gaussian model M is defined according to the standard deviation $\sigma > 0$. The window size $[w, w]$ can then be determined as $w \approx 5\times\sigma$, and the Gaussian model M is defined as the $w\times w$ truncation from the Gaussian kernel $G(x, y)$. We then calculate the norm of the discriminant vector $\|v_i\|_2=\sqrt{v_i^T v_i}$, and map it into the corresponding discriminant image I_i. The filter M is used to smooth the discriminant image

$$I_i'(x,y)=I(x,y)*M(x,y). \tag{18}$$

$I_i'(x,y)$ is transformed into a high dimensional vector v_i' by concatenating the rows of $I_i'(x,y)$ together. Finally we normalize v_i' using the norm of v_i

$$v_i''=\frac{\|v_i\|_2}{\sqrt{v_i'^T v_i'}}v_i', \tag{19}$$

and obtain the post-processed discriminant vector v_2''.

Compared with other LDA techniques, the post-processed LDA method has some potential advantages, such as directness, two dimensionality, and complementarity. First, post-processed LDA is designed to directly modify the discriminant vectors. Other LDA techniques, such as EFM, usually adopt the strategy to define the within-class scatter matrix in PCA subspace. Second, when applied to image recognition task, post-processed LDA maps a discriminant vector into a two-dimensional image, and thus can use two-dimensional image processing techniques to alter the appearance of the discriminant vector. Third, post-processing can be used as a complementary approach to combine with other LDA techniques, such as enhanced Fisher model, and completer Fisher discriminant framework.

3.1.2 Post-processed Enhanced Fisher Model

The Enhanced Fisher Model (EFM) method is based on the PCA plus LDA framework where PCA is used to alleviate the over-fitting problem and to improve the generalization

performance (Liu & Wechsler, 1998; Liu & Wechsler, 2002). In (Wang & Tang, 2004), Wang and Tang present another insight to understand EFM by modeling face difference with intrinsic difference, transform difference, and noise, where the PCA transform is used to significantly reduce noise, and the subsequent LDA step is used to separate intrinsic difference from transform difference.

In EFM, each image should be previously mapped into a one-dimensional vector by concatenating the rows of the original image. Let $\mathbf{X} = \{\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \cdots, \mathbf{x}_{N_1}^{(1)}, \cdots, \mathbf{x}_j^{(i)}, \cdots, \mathbf{x}_{N_C}^{(C)}\}$ be a training set with N_i image vectors for class i. The number of class is C, and $\mathbf{x}_j^{(i)}$ denotes the jth image vector of class i. The total covariance matrix $\mathbf{S}_t$ of PCA is then defined as

$$\mathbf{S}_t = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(\mathbf{x}_j^{(i)} - \overline{\mathbf{x}})(\mathbf{x}_j^{(i)} - \overline{\mathbf{x}})^T , \tag{20}$$

where $\overline{\mathbf{x}}$ is the mean vectors of all training images, and $N = \sum_{i=1}^{C} N_i$ is the total number of training images. The PCA projector $\mathbf{T}_{pca} = [\varphi_1, \varphi_2, \cdots, \varphi_{d_{\text{PCA}}}]$ can be obtained by calculating the eigenvalues and vectors of the total scatter matrix $\mathbf{S}_t$, where φ_k is the kth eigenvector corresponding to the kth largest eigenvalue of $\mathbf{S}_t$, and d_{PCA} denotes the PCA dimension for the EFM method.

The between-class scatter matrix $\mathbf{S}_b$ and the within-class scatter matrix $\mathbf{S}_w$ are defined as

$$\mathbf{S}_b = \frac{1}{N}\sum_{i=1}^{C} N_i(\overline{\mathbf{x}^{(i)}} - \overline{\mathbf{x}})(\overline{\mathbf{x}^{(i)}} - \overline{\mathbf{x}})^T , \tag{21}$$

$$\mathbf{S}_w = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(\mathbf{x}_j^{(i)} - \overline{\mathbf{x}^{(i)}})(\mathbf{x}_j^{(i)} - \overline{\mathbf{x}^{(i)}})^T , \tag{22}$$

where $\overline{\mathbf{x}}^{(i)}$ is mean vector of class i. With PCA projector $\mathbf{T}_{pca}$, we map $\mathbf{S}_b$ and $\mathbf{S}_w$ to the PCA subspace,

$$\breve{\mathbf{S}}_b = \mathbf{T}_{pca}^T \mathbf{S}_b \mathbf{T}_{pca} \text{ and } \breve{\mathbf{S}}_w = \mathbf{T}_{pca}^T \mathbf{S}_w \mathbf{T}_{pca} . \tag{23}$$

PCA projection can eliminate the singularity of the within-class scatter matrix. Thus the optimal discriminant vectors can be calculated by maximizing the Fisher's criterion

$$J_F(w) = \frac{w^T \breve{\mathbf{S}}_b w}{w^T \breve{\mathbf{S}}_w w} . \tag{24}$$

The discriminant vectors can be obtained by calculate the first d_{LDA} generalized eigenvectors $[w_1, w_2, \cdots, w_{d_{\text{LDA}}}]$ and the corresponding eigenvalues $[\lambda_1, \lambda_2, \cdots, \lambda_{d_{\text{LDA}}}]$ of $\breve{\mathbf{S}}_b$ and $\breve{\mathbf{S}}_w$. Given an image vector $\mathbf{x}$, the discriminant feature vector $\mathbf{z}^S$ is defined as

$$\mathbf{z}^S = \mathbf{U}_S^T \mathbf{T}_{SPCA}^T \mathbf{x} , \tag{25}$$

where $\mathbf{U}_S = [w_1, w_2, \cdots, w_{d_{\text{LDA}}}]$ is the subspace LDA projector.

PEFM Algorithm

Step 1. Compute $\mathbf{S}_t$, $\mathbf{S}_b$, and $\mathbf{S}_w$, and calculate the first d_{PCA} eigenvectors $\mathbf{T}_{pca} = [\varphi_1, \varphi_2, \cdots, \varphi_{d_{\text{PCA}}}]$ of $\mathbf{S}_t$, then modify $\mathbf{S}_b$ and $\mathbf{S}_w$ by $\tilde{\mathbf{S}}_b = \mathbf{T}_{pca}^T \mathbf{S}_b \mathbf{T}_{pca}$ and $\tilde{\mathbf{S}}_w = \mathbf{T}_{pca}^T \mathbf{S}_w \mathbf{T}_{pca}$.

Step 2. Calculate the first d_{LDA} generalized eigenvectors $\mathbf{U}_S = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{d_{\text{LDA}}}]$ of $\tilde{\mathbf{S}}_b$ and $\tilde{\mathbf{S}}_w$, and compute $\mathbf{T}_{LDA} = \mathbf{T}_{PCA}\mathbf{U}_S = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{d_{\text{LDA}}}]$.

Step 3. Using 2D-Gaussian filter $\mathbf{M}_g$, regularize each discriminant vector $\mathbf{v}_i$ to $\mathbf{v}_i''$, and construct the LDA projector $\mathbf{T}_{PEFM} = [\mathbf{v}_1'', \mathbf{v}_2'', \cdots, \mathbf{v}_{d_{\text{LDA}}}'']$.

Fig. 5. PEFM Algorithm

With the post-processing approach, we present the implementation of the post-processed RFM method (PEFM), where the main steps are illustrated in Fig. 5.

3.2 Relation between the post-processing approach and image Euclidean distance

In (Wang et al., 2005), Wang et al presented an image Euclidean distance (IMED) method, where a 2D-Gaussian function is used to model the effect of neighbor pixels. Compared with traditional Euclidean distance, IMED can be easily embedded with some popular image feature extraction and classification methods and reported a consistent performance improvement. In this section, we will demonstrate that the IMED method actually is equivalent to the post-processing approach.

Different from traditional Euclidean distance, the computation of image Euclidean distance take into account the spatial relationships of pixels

$$d_{IME}^2(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2\pi}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k}^{n}\sum_{l}^{n}[\mathbf{X}_1(i,j) - \mathbf{X}_2(i,j)][\mathbf{X}_1(k,l) - \mathbf{X}_2(k,l)]\exp\{-\frac{(i-k)^2 + (j-l)^2}{2}\}, \quad (26)$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are two are two $m \times n$ images, and $\mathbf{X}_1(i, j)$ represent the gray value of the (i, j) pixel of image $\mathbf{X}_1$.

Traditional Euclidean distance can be easily rewritten using its inner product representation

$$d_E^2(\mathbf{X}_1, \mathbf{X}_2) = \langle \mathbf{X}_1, \mathbf{X}_1 \rangle + \langle \mathbf{X}_2, \mathbf{X}_2 \rangle - 2\langle \mathbf{X}_1, \mathbf{X}_2 \rangle. \quad (27)$$

Similarly, by defining image inner product (IMIP) as

$$\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_{IMIP} = \frac{1}{2\pi}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k}^{n}\sum_{l}^{n}\exp\{-\frac{(i-k)^2 + (j-l)^2}{2}\}\mathbf{X}_1(i,j)\mathbf{X}_2(k,l), \quad (28)$$

image Euclidean distance can be re-formalized as

$$d_{IME}^2(\mathbf{X}_1, \mathbf{X}_2) = \langle \mathbf{X}_1, \mathbf{X}_1 \rangle_{IMIP} + \langle \mathbf{X}_2, \mathbf{X}_2 \rangle_{IMIP} - 2\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_{IMIP}. \quad (29)$$

Different from traditional inner product, IMIP not only consider the product of two corresponding pixels, but also consider the effect of the spatial relationship between neighbor pixels, and thus is more robust against small degree of variations in translation, rotation and deformation. With the introduction of IMED and IMIP, we can conveniently embed them into many popular image feature extraction and classification approaches, such as PCA, LDA, k-nearest neighbor, and support vector machine.

According to the separability property, the definition of IMIP can be formalized to

$$\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_{IMIP} = \frac{1}{2\pi} \sum_{i=1}^{m} \sum_{k=1}^{m} \exp\{-\frac{(i-k)^2}{2}\} \sum_{j=1}^{n} \sum_{l=1}^{n} \exp\{-\frac{(j-l)^2}{2}\} \mathbf{X}_1(i,j) \mathbf{X}_2(k,l) . \quad (30)$$

Let

$$\mathbf{X}'_1(i,l) = \frac{1}{\sqrt{2\pi}} \sum_{j=1}^{n} \exp\{-\frac{(j-l)^2}{2}\} \mathbf{X}_1(i,j) , \quad (31)$$

$$\mathbf{X}''_1(k,l) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{m} \exp\{-\frac{(i-k)^2}{2}\} \mathbf{X}'_1(i,l) , \quad (32)$$

IMIP can then be represented as

$$\langle \mathbf{X}_1, \mathbf{X}_2 \rangle_{IMIP} = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{X}''_1(i,j) \mathbf{X}_2(i,j) = \langle \mathbf{X}''_1, \mathbf{X}_2 \rangle . \quad (33)$$

From Eq. (31) and (32),

$$\mathbf{X}''_1(k,l) = \frac{1}{2\pi} \sum_{j=1}^{n} \sum_{i=1}^{m} \exp\{-\frac{(i-k)^2 + (j-l)^2}{2}\} \mathbf{X}_1(i,j) . \quad (34)$$

In the spatial domain, the definition of two-dimensional linear convolution is

$$g(i,j) = f(i,j) * h(i,j) = \sum_{k=1}^{m} \sum_{l=1}^{n} f(k,l) h(i-k, j-l) . \quad (35)$$

where $f(k,l)$ denotes the original image, $h(i,j)$ denotes the convolution kernel, and $g(i,j)$ denotes the convolution result. Clearly, by defining the convolution kernel

$$h(i,j) = \frac{1}{2\pi} \exp\{-\frac{i^2 + j^2}{2}\} , \quad (36)$$

we can show the equivalence of IMIP and the post-processing approach. IMIP actually is the post-processing approach with the standard deviation $\sigma = 1$ and without the normalization step. The post-processing approach, in fact, can be regarded as a generalization of the normalized IMIP method without the constraint on the value of the standard deviation.

3.3 Performance evaluation of PEFM

In this section, we use the FERET face database to evaluate the efficiency of the PEFM method over the original EFM method, to verify the equivalence of IMED and the post-processing approach, and to evaluate the influence of the normalization step.

In our experiment, we adopt the same experimental setup as described in Section 2.3. The nearest neighbor classifier is used to match probe images and gallery images, and the averaged recognition rate (ARR) is adopted by calculating the mean value of recognition rates across 10 runs.

For PEFM, there are three parameters, the PCA dimension d_{PCA}, the LDA dimension d_{LDA}, and the standard deviation σ, to be determined. However, it is very difficult to determine these three parameters at the same time. Previous work on the FERET subset has shown that the maximum recognition accuracy could be obtained with the LDA dimension $d_{\mathrm{LDA}} \approx 20$. With standard deviation $\sigma \approx 1.5$, the noise in the discriminant vector would be significantly reduced. So we investigate the effect of the PCA dimension d_{PCA} with $d_{\mathrm{LDA}} = 20$ and $\sigma = 1.5$,. As the PCA dimension d_{PCA} (>100) increases, PEFM will be distinctly superior to EFM in terms of recognition accuracy. Besides, the PCA dimension has a much less effect on the recognition accuracy of PEFM, whereas that of EFM deteriorates greatly with increasing of d_{PCA}. From Fig. 6(a), we can determine the PCA dimension of PEFM, $d_{\mathrm{PCA}} = 100$.
After determining d_{PCA}, we study the recognition accuracy over the variation of the LDA dimension d_{LDA} with $d_{\mathrm{PCA}} = 100$ and $\sigma = 1.5$, as depicted in Fig. 6(b). The maximum average recognition rate of PEFM is obtained with the LDA dimension $d_{\mathrm{LDA}} = 24$. Then we explore the recognition rate vs. the variation of σ with $d_{\mathrm{PCA}} = 100$ and $d_{\mathrm{LDA}} = 24$, as shown in Fig. 6(c). The maximum average recognition rate, 87.34%, is obtained using PEFM with $d_{\mathrm{PCA}} = 100$, $d_{\mathrm{LDA}} = 24$ and $\sigma = 1.5$, which is higher than 84.54%, that of EFM.

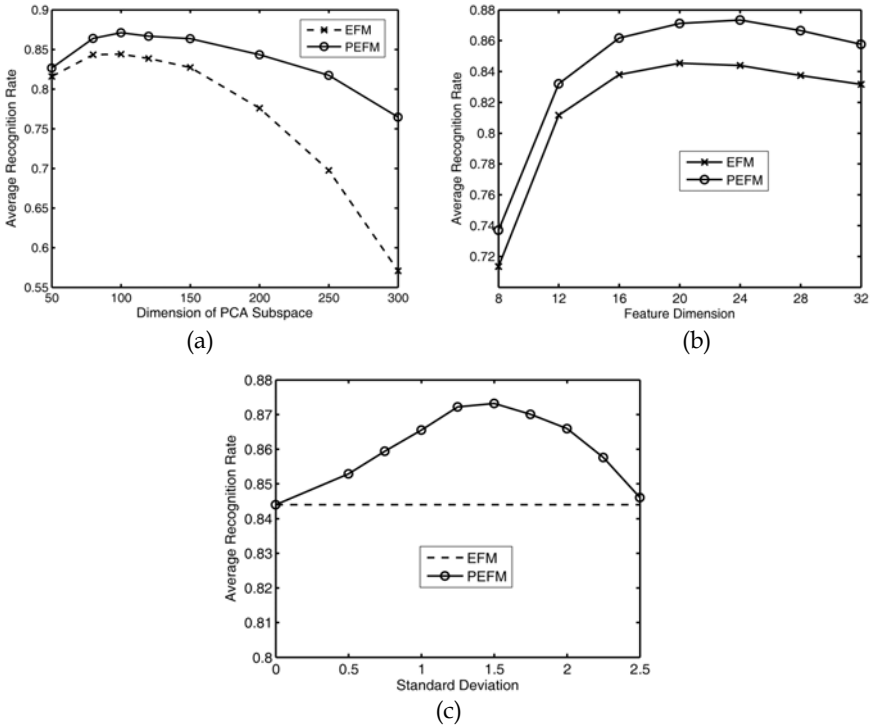


Fig. 6. Illustration of the recognition accuracy over the variation of the PEFM parameters on the FERET subset. (a) Recognition accuracy vs. the PCA dimension. (b) Recognition accuracy vs. the LDA dimension. (c) Recognition accuracy vs. the standard variation

Next we compare the recognition performance of IMED-embedded EFM and PEFM without normalization. Fig. 7 shows the recognition rates of these two methods over different feature

dimensions. From Fig. 7, we can see that, the performance difference between these two approaches is very small, and PEFM without normalization only achieve a little higher recognition rate than IMED-embedded EFM. The small performance difference, however, may be explained by that, for PEFM, IMED is only embedded in the testing stage. This indicates that, when IMED is embedded in enhanced Fisher model (EFM), it would be better to embed IMED only in the testing stage rather than to embed IMED in both the training and the testing stage.

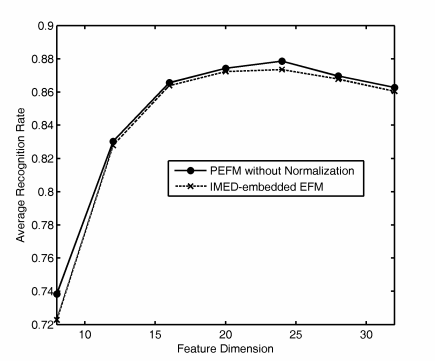


Fig. 7. Recognition rates of IMED-embedded EFM and PEFM without normalization

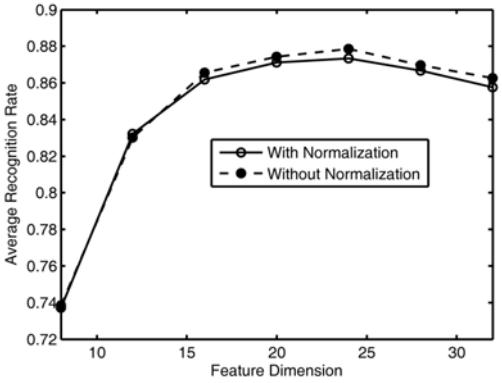


Fig. 8. Illustration of the recognition rates of PEFM with and without the normalization step over different feature dimensions.

Finally, we investigate the influence of the normalization step. Fig. 8 shows the recognition rates of PEFM with and without the normalization step over different LDA dimensions. From Fig. 8, we can see that, the normalization step in PEFM actually has a little effect on the recognition performance when applied to face recognition.

4. Robust recognition by iterated reweighted fitting of eigenfaces

A real face recognition system should capture, detect and recognize facial image automatically, making it inevitable that facial images will sometimes be noisy, partially occluded, or inaccurately located. The capture and communication of facial image itself may

introduce noise; some accessories will cause the occlusion of a facial image, for example a scarf may occlude a facial image; and facial images usually should be normalized by locations of landmarks but these locations may be inaccurate and inconsistent. Because all these three factors are inevitable, the development of face recognition system should always address the robust recognition of noisy, partially occluded, or inaccurate located image.

4.1 Iterated reweighted fitting of eigenfaces

Zhao et al. (2003) showed that, in face recognition, the appearance-based methods (e.g., PCA and LDA) are robust in the presence of low levels of small noise or occlusion. However, if the degree of occlusion further increases, the recognition performance would deteriorate severely (Martinez, 2002). Analogous to partial occlusion, the further increase of noise would also cause an immediate decrease in recognition performance.

To address the partial occlusion problem, Martinez proposed a local probabilistic approach where each face image is divided into six local areas, and each local area is then projected into its eigenspace in the recognition stage (Martinez, 2002). Local probabilistic approach, however, cannot be used to weaken the unfavorable effect of noise because noise is always globally distributed. Besides, local probabilistic approach, which divides an image into a number of parts, also neglects the global correlation of the face image.

Robust estimation (McLachlan & Krishnan, 1997; Isao & Eguchi, 2004) and robust appearance-based methods can be used to solve the noise and partial occlusion problems. For example, iterated reweighted fitting of Eigenfaces (IRF-Eigenfaces), a robust estimation of the coefficients of Eigenfaces, can address this by first defining an objective function $J(\mathbf{y})$ and then using the Expectation Maximization (EM) algorithm to compute the feature vector $\mathbf{y}$ by minimizing $J(\mathbf{y})$. The following presents the main steps in IRF-Eigenfaces:

1. Define the objective function. Given a set of Eigenfaces $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_d]$, IRF-Eigenfaces calculate the feature vector $\mathbf{y}$ of image $\mathbf{x}$ by minimizing the objective function

$$J(\mathbf{y}) = \sum_{i=1}^{mn} \Psi\left((x_i - \mathbf{W}_i\mathbf{y})^2\right), \tag{37}$$

where the function $\Psi(z)$ could be defined as (Isao & Eguchi, 2004)

$$\Psi(z) = \log\frac{1}{1+\exp(-\beta(z-\eta))}, \tag{38}$$

where the inverse temperature β and saturation value η are two tuning parameters.

2. Calculate feature vector $\mathbf{y}$ by iteratively performing the next two steps until the value of $\mathbf{y}^{(t)}$ converges or t arrives at the pre-determined threshold $t_{\max}$.:

E-Step. Given $\mathbf{y}^{(t)}$, update the weighted vector $\boldsymbol{\omega}^{(t)} = [\omega_1^{(t)}, \omega_2^{(t)}, \cdots, \omega_{mn}^{(t)}]$ by

$$\omega_i^{(t)} = \varphi\left(z_i^{(t)}\right) = \frac{\exp\left\{-\beta(z_i^{(t)} - \eta)\right\}}{1+\exp\left\{-\beta(z_i^{(t)} - \eta)\right\}}$$

$$z_i^{(t)} = (x_i - \mathbf{W}_i\mathbf{y}^{(t)})^2$$

M-Step. Given the weighted vector $\boldsymbol{\omega}^{(t)}$, update $\mathbf{y}^{(t+1)}$ by

$$\mathbf{y}^{(t+1)} = \left(\sum_{i=1}^{mn} \omega_i^{(t)} \mathbf{W}_i^T \mathbf{W}_i \right)^{-1} \sum_{i}^{mn} \omega_i^{(t)} \mathbf{W}_i^T x_i$$

If we define the function $\Psi(z)$ a strictly concave function in z, $\Psi''(z) < 0$, we can theoretically guarantee that $\mathbf{y}^{(t)}$ will converge to a local optimal feature vector $\mathbf{y}$.

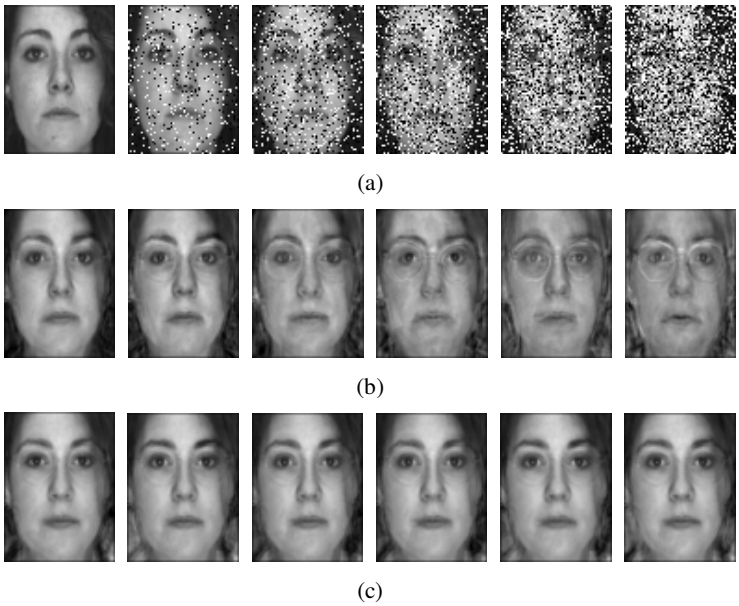
4.2 Evaluation on IRF-eigenfaces

We use the AR face database to evaluate the performance of the IRF-Eigenfaces method against noise and partial occlusion, and compare the recognition rate of IRF-Eigenfaces with those of Eigenfaces and the local probabilistic approach. The AR face database contains over 4,000 color frontal facial images corresponding to 126 people (70 men and 56 women) (Martinez & Benavente, 1998).There are 26 different images of each person and these were captured in two sessions separated by two weeks. In our experiments, we only use a subset of 720 images of 120 persons. There are six images of each person. Three images were captured in the first session (one neutral, one with sunglasses, and one with a scarf) and the remaining three images were captured in the second session (one neutral, one with sunglasses, and one with a scarf), as shown in Fig. 9. In our experiments, all the images were cropped according to the location of their eyes and mouths, and the 120 neutral images in the first session were used for training Eigenfaces. We set the number of principal components d_{PCA}=100 and use the whitened cosine distance measure (Phillips et al., 2005).

Fig. 9. Six images of one person in the AR database. The images (a) through (c) were captured during one session and the images (d) through (f) at a different session.

IRF-Eigenfaces is more robust when it comes to reconstruct noisy and occluded facial images. The quality of reconstructed facial images using IRF-Eigenfaces is also consistently better than those using Eigenfaces, as shown in Fig. 10. IRF-Eigenfaces also has a robust reconstruction performance for the partially occluded facial images, as shown in Fig. 11.
Using all the neutral images in the second session as a test set, we investigate the recognition performance of IRF-Eigenfaces against different degree of noise. Fig. 12 shows an original facial image and the same image after the addition of various amounts of salt and pepper

noise. The largest degree of noise in our test is 50%, which is a seriously contaminated example. We then present the recognition rates of Eigenfaces and IRF-Eigenfaces against different degrees of noise contamination, as shown in Fig. 13. The addition of 50% salt and pepper noise caused the recognition rate of Eigenfaces to fall from 81.67% to 36.67%, but the recognition rate of IRF-Eigenfaces remained unchanged (95.83%). Because the recognition rate is robust against variation of noise, we can validate the robustness of IRF-Eigenfaces against noise.

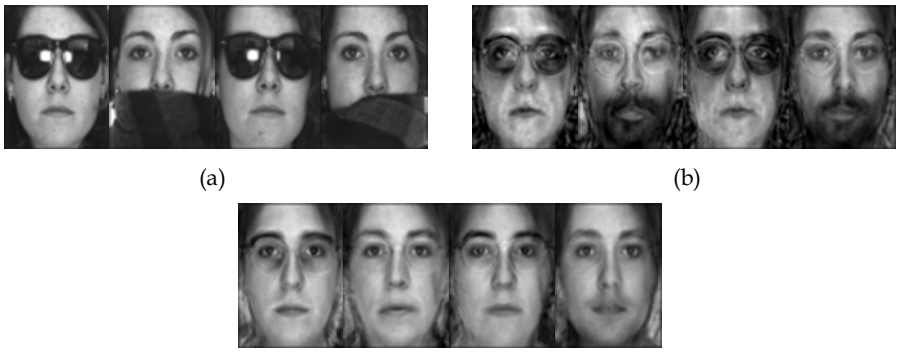


(a)

(b)

(c)

Fig. 10. Reconstruction of images with different degree of salt and pepper noise: (a) original image; (b) reconstructed images by Eigenfaces; (c) reconstructed images by IRF-Eigenfaces.



(a) (b)

(c)

Fig. 11. Reconstructed partially occluded images: (a) original images; (b) reconstructed images by Eigenfaces; (c) reconstructed images by IRF-Eigenfaces.

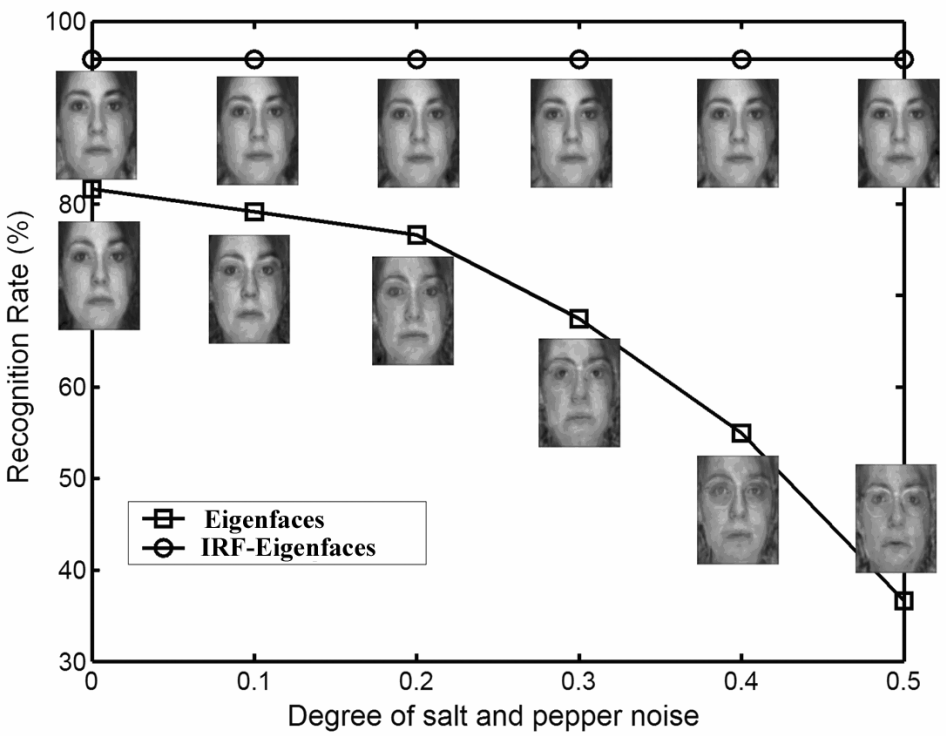


Fig. 13. The effect of salt and pepper noise on the recognition performance

Using the images with sunglasses or scarves, we tested the influence of partial occlusion on the recognition performance of IRF-Eigenfaces. Table 5 lists the recognition rates of Eigenfaces, IRF-Eigenfaces and the local probabilistic approach in recognizing faces partially occluded with either sunglasses or a scarf. The recognition rate of IRF-Eigenfaces in both the first and second sessions is much higher than that of the other two methods, Eigenfaces and the local probabilistic approach.

Methods	Session 1		Session 2	
	Sunglasses	Scarf	Sunglasses	Scarf
Eigenfaces (%)	40	35	26.67	25
LocProb (%)	80	82	54	48
IRF-Eigenfaces (%)	87.50	91.67	82.50	84.17

Table 5. Recognition performance of three face recognition methods on the AR database

Another interesting point to be noted from Table 5 is that IRF-Eigenfaces is also more robust against the variation of ageing time. Eigenfaces' recognition rate in the second session (25.83%) is much lower than in the first session (37.5%). The local probabilistic approach's recognition rate in the second session (51%) is much lower than in the first session (81%). But IRF-Eigenfaces' recognition rate in the second session (83.34%) is only slightly lower than in the first session (89.58%). There are two reasons for the robustness of IRF-Eigenfaces

against ageing time. First, IRF-Eigenfaces uses the whitened cosine distance, which can reduce the adverse effect of global illumination change of the facial image. Second, IRF-Eigenfaces, which is robust to partial occlusion, is also robust to some facial change, such as the presence of a beard. Compare Fig. 14(a), showing a neutral face captured in the first session, with Fig. 14(b), showing a face with sunglasses captured in the second session. The image in Fig. 14(b), captured in the second session, has a heavier beard. Fig. 14(c) shows the image in Fig. 14(b) reconstructed using IRF-Eigenfaces. IRF-Eigenfaces can detect parts of the beard as a partial occlusion and thus its reconstructed image is more consistent with Fig. 14(a).

(a) (b) (c)

Fig. 14. The reconstruction performance of IRF-Eigenfaces for partially occluded face in the second session. (a) neutral face in the first session; (b) face with sunglasses in the second session; (c) the reconstructed image of (b) using IRF-Eigenfaces.

7. Summary

In this chapter, we introduce several recently developed subspace-based face recognition methods in addressing three problems, singularity, regularization, and robustness. To address the singularity problem, we present a fast feature extraction technique, Bi-Directional PCA plus LDA (BDPCA+LDA), which performs LDA in the BDPCA subspace. Compared with the PCA+LDA framework, BDPCA+LDA has a number of significant advantages. First, BDPCA+LDA needs less computational requirement in both the training and the testing phases. Second, BDPCA+LDA needs less memory requirement because its projector is much smaller than that of PCA+LDA. Third, BDPCA+LDA has a higher recognition accuracy over PCA+LDA.

To alleviate the over-fitting to the training set, this chapter suggests a post-processing approach on discriminant vectors, and demonstrates the internal relationship between the post-processing approach and IMED. Experimental results indicate that, the post-processing approach is effective in improving the recognition rate of the LDA-based approaches. When IMED is embedded in enhanced Fisher model, it would be better to embed IMED only in the testing stage.

To improve the robustness of subspace method over noise and partial occlusion, this chapter further presents an iteratively reweighted fitting of the Eigenfaces method (IRF-Eigenfaces). Despite the success of IRF-Eigenfaces in recognizing noisy and partially occluded facial images, it is still very necessary to further study this issue by investigating the robustness against inaccurate fiducial point location, illumination, and ageing in one uniform framework.

8. Acknowledgement

The work is partially supported by the 863 fund under No. 2006AA01Z308, the development program for outstanding young teachers in Harbin Institute of Technology under No.

HITQNJS.2008.049. The author would like to thank Jian Yang and Yong Xu for their constructive suggestions to our research work.

9. References

Baeka, J. & Kimb, M. (2004) Face recognition using partial least squares components. *Pattern Recognition*, Vol. 37, 1303-1306

Bartlett, M.S.; Movellan, J.R. & Sejnowski, T.J. (2002) Face Recognition by Independent Component Analysis. *IEEE Trans. Neural Network*, Vol. 13, No. 6, 1450-1464

Belhumeur, P. N.; Hespanha, J. P. & Kriegman, D. J. (1997) Eigenfaces vs Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.20, No.7, 711-720

Cevikalp, H.; Neamtu, M.; Wilkes, M. & Barkana, A. (2005) Discriminative common vectors for face recognition. *IEEE Trans Pattern Analysis and Machine Intelligence*, Vol. 27, No. 1, 4-13

Fukunaga, K. (1990) *Introduction to Statistical Pattern Recognition, Second Edition*, Academic Press, New York

He, X.; Yan, S.; Hu, Y.; Niyogi, P. & Zhang, H.-J. (2005) Face Recognition Using Laplacianfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 3, 328-340

Isao, I. & Eguchi, S. (2004) Robust principal component analysis with adaptive selection for tuning parameters. *Journal of Machine Learning Research*, Vol. 5, 453-471

Jolliffe, I.T. (2002) *Principal Component Analysis, Second Edition*, Springer, New York

Kanade, T. (1973) *Computer recognition of human faces*. Birkhauser, Basel, Switzerland, and Stuttgart, Germany

Lathauwer, L.; Moor, B. & Vndewalle, J. (2000) A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Application*, Vol. 21, No. 4, 1233-1278

Liu, C. (2004) Gabor-based kernel PCA with fractional power polynomial models for face recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, No. 5, 572-581

Liu, C. & Wechsler, H. (1998) Enhanced Fisher linear discriminant models for face recognition. *Proc. 14th Int'l Conf. Pattern Recognition*, Vol. 2, pp. 1368-1372

Liu, C. & Wechsler, H. (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Processing*, Vol. 10, No. 4, 467-476

Liu, K.; Cheng, Y.-Q. & Yang, J.-Y. (1993) Algebraic feature extraction for image recognition based on an optimal discriminant criterion. *Pattern Recognition*, Vol. 26, No. 6, 903-911

Lu, J.; Plataniotis, K.N. & Venetsanopoulos, A.N. (2003) Face recognition using kernel direct discriminant analysis algorithms. *IEEE Trans. Neural Networks*, Vol. 14, No. 1, 117-126

Martinez, A.M. (2002) Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 6, 748-763

Martinez, A.M. & Benavente, R. (1998) *The AR Face Database*. CVC Technical Report #24

McLachlan, G.J. & Krishnan, T. (1997) *The EM algorithm and extensions*, John Wiley & Sons, New York

Philips, P.J.; (1998) The Facial Recognition Technology (FERET) Database, http://www.itl. nist.gov/iad /humanid/feret.

Phillips, P.J.; Flynn, P.J.; Scruggs, T.; Bowyer, K.W.; Chang, J.; Hoffman, K.; Marques, J.; Min, J. & Worek, W. (2005) Overview of the face recognition grand challenge. *Proc IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 947-954

Philips, P.J.; Moon, H.; Rizvi, S.A. & Rauss, P.J. (2000) The FERET evaluation methodology for face-recognition algorithms,"*IEEE Trans Pattern Analysis and Machine Intelligence*, Vol. 22, No. 10, 1090-1104

Roweis, S.T. & Saul, L.K. (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science*, Vol. 290, 2323-2326

Swets, D.L. & Weng, J. (1996) Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8, 831-836

Tao, D.; Li, X.; Hu, W.; Maybank, S.J. & Wu, X. (2005) Supervised tensor learning. *IEEE Int'l Conf. on Data Mining*, pp. 450-457

Tenenbaum, J.B.; de Silva, V.; & Langford, J.C. (2000) A global geometric framework for nonlinear dimensionality reduction. *Science*, Vol. 290, 2319-2323

Timo, A.; Abdenour, H. & Matti, P. (2004) Face Recognition with Local Binary Patterns. *Proceeding of European Conference on Computer Vision*, pp. 469-481

Torkkola, K. (2001) Linear Discriminant Analysis in Document Classification. *Proc. IEEE ICDM Workshop Text Mining*

Turk, M. & Pentland, A. (1991) Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, Vol.3, No.1, 71-86

Wang, L.; Zhang, Y. & Feng, J. (2005) On the Euclidean distance of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, 1334-1339

Wang, K.; Zuo, W. & Zhang, D. (2005) Post-processing on LDA's discriminant vectors for facial feature extraction. *Proc. 5th Int'l Conf. Audio- and Video-based Biometric Person Authentication*, pp. 346-354

Wang, X. & Tang, X. (2004) A unified framework for subspace face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, 1222-1228

Wiskott, L.; Fellous, J. M.; Kruger, N. & Malsburg, C. (1997) Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No. 7, 775-779

Yan, S.; Xu, D.; Yang, Q.; Zhang, L.; Tang, X. & Zhang, H.-J. (2007) Multilinear discriminant analysis for face recognition. *IEEE Trans. Image Processing*, Vol. 16, No. 1, 212-220

Yang, J. & Yang, J.Y. (2003) Why can LDA be performed in PCA transformed space. *Pattern Recognition*, Vol. 36, No. 2, 563-566

Yang, J.; Zhang, D.; Frangi, A.F. & Yang, J.-Y. (2004) Two-Dimensional PCA: a New Approach to Face Representation and Recognition. *IEEE Trans Pattern Analysis and Machine Intelligence*, Vol. 26, No. 1, 131-137

Yang, J.; Zhang, D.; Xu, Y. & Yang, J.Y.(2005a) Two-dimensional discriminant transform for face recognition. *Pattern Recognition*, Vol. 38, No. 7, 1125-1129

Yang, J.; Zhang, D.; Yang, J.-Y.; Zhong, J. & Frangi, A.F. (2005b) KPCA plus LDA: a Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 2, 230-244

Yang, M.H. (2002) Kernel Eigenfaces vs. kernel Fisherfaces: face recognition using kernel methods. *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (RGR'02)*, pp. 215-220

Ye, J. (2004) Generalized Low Rank Approximations of Matrices. *The Twenty-First International Conference on Machine Learning*

Ye, J. (2005) Generalized low rank approximations of matrices. *Machine Learning*, Vol. 61, No. 1-3, 167-191

Yu, H. & Yang, J.(2001) A direct LDA algorithm for high-dimensional data—with application to face recognition. *Pattern Recognition*, Vol. 34, No. 10, 2067-2070

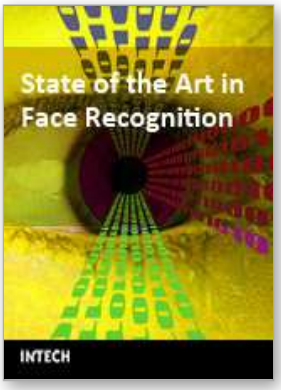
Zafeiriou, S.; Tefas, A.; Buciu, I. & Pitas, I. (2006) Exploiting Discriminant Information in Nonnegative Matrix Factorization With Application to Frontal Face Verification. *IEEE Trans. Neural Networks*, Vol. 17, No. 3, 683-695

Zhao, W.; Chellappa, R.; Phillips, P.J. & Rosenfeld, A.(2003) Face recognition: a literature survey. *ACM Computing Surveys*, Vol. 35, No. 4, 399-458

Zuo, W.; Wang, K. & Zhang, D.(2005a) Bi-directional PCA with assembled matrix distance metric. *International Conference on Image Processing*, pp. 958-961

Zuo, W.; Wang, K.; Zhang, D. & Yang, J. (2005b) Regularization of LDA for face recognition: a post-processing approach. *Proc. 2th Int'l Workshop on Analysis and Modeling of Face and Gesture*, pp. 377-391

Zuo, W.; Zhang, D.; Yang, J. & Wang, K. (2006) BDPCA plus LDA: a novel fast feature extraction technique for face recognition. *IEEE Trans. Systems, Man, and Cybernetics, Part B*, Vol. 36, No. 4, 946-953



State of the Art in Face Recognition

Edited by Julio Ponce and Adem Karahoca

ISBN 978-3-902613-42-4
Hard cover, 436 pages
Publisher I-Tech Education and Publishing
Published online 01, January, 2009
Published in print edition January, 2009

Notwithstanding the tremendous effort to solve the face recognition problem, it is not possible yet to design a face recognition system with a potential close to human performance. New computer vision and pattern recognition approaches need to be investigated. Even new knowledge and perspectives from different fields like, psychology and neuroscience must be incorporated into the current field of face recognition to design a robust face recognition system. Indeed, many more efforts are required to end up with a human like face recognition system. This book tries to make an effort to reduce the gap between the previous face recognition research state and the future state.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Wangmeng Zuo, Kuanquan Wang and Hongzhi Zhang (2009). Subspace Methods for Face Recognition: Singularity, Regularization, and Robustness, State of the Art in Face Recognition, Julio Ponce and Adem Karahoca (Ed.), ISBN: 978-3-902613-42-4, InTech, Available from: http://www.intechopen.com/books/state_of_the_art_in_face_recognition/subspace_methods_for_face_recognition__singularity__regularization__and_robustness

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821