Applied Mathematics and Computation 205 (2008) 726–734

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Face recognition using FLDA with single training image per person

Quan-xue Gao [a,b], Lei Zhang [a,*], David Zhang [a]

[a] Biometrics Research Center, Department of Computing, The Kong Kong Polytechnic University, Hong Kong, China
[b] School of Telecommunications Engineering, XIDIAN University, 710071, Xi'an, China

ARTICLE INFO

Keywords:
Face recognition
Fisher linear discriminant analysis (FLDA)
Single training image per person
Singular value decomposition (SVD)

ABSTRACT

Fisher linear discriminant analysis (FLDA) has been widely used for feature extraction in face recognition. However, it cannot be used when each object has only one training sample because the intra-class variations cannot be statistically measured in this case. In this paper, a novel method is proposed to solve this problem by evaluating the within-class scatter matrix from the available single training image. By using singular value decomposition (SVD), we decompose the face image into two complementary parts: a smooth general appearance image and a difference image. The later is used to approximately evaluate the within-class scatter matrix and thus the FLDA can be applied to extract the discriminant face features. Experimental results show that the proposed method is efficient and it can achieve higher recognition accuracy than many existing schemes.

1. Introduction

Face recognition has been extensively studied for many years and it is still attracting much attention because of its big potential in security, surveillance and human-computer intelligent interaction, etc. A key issue in face recognition is to find sufficient and discriminative features for face representation. Many approaches have been proposed and subspace analysis method (SAM) has become one of the most popular methods. SAM seeks for a set of basis vectors according to some criteria and extracts the features by projecting face images onto the subspace spanned by those basis vectors [1–3].

Principal component analysis (PCA), which tries to find a set of optimal orthogonal bases in the sense of minimum mean square error, and Fisher linear discriminant analysis (FLDA), which tries to find a set of optimal projection vectors by maximizing the ratio between the determinants of the between-class and the within-class scatter matrices of the training samples, are the two most representative methods in SAM. By first applying PCA to face recognition, Kirby and Sirovich [4] found that a face image could be reconstructed approximately as a weighted sum of a small collection of basis face images plus a mean face image. Based on this work, Turk and Pentland [5] developed the well-known Eigenface method. Since then, PCA has been extensively investigated and many PCA-based algorithms have been developed [6,7]. Although PCA enables sufficient reconstruction, it may not be optimal for classification because its optimality is in the sense of minimum mean square error [8]. To improve the classification performance, we need to combine further this optimal representation criterion with some discrimination criterion.

FLDA was proposed to this end and it has become a widely used discriminant criterion in face recognition. FLDA aims to find a set of projection bases, which could be able to separate the samples of different classes as far as possible and compress the same class samples as compact as possible. By applying it to face recognition, Belhumeur et al. [9] proposed the well-known Fisherface algorithm. Similar methods were proposed by Swets et al. [10] and by Etemad et al. [11]. Fisherface method first employs PCA to obtain a low dimensional space and then implements FLDA to extract the discriminant face features.

* Corresponding author.
E-mail address: cslzhang@comp.polyu.edu.hk (L. Zhang).

0096-3003/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2008.05.019

Comparative studies between FLDA and PCA on face recognition were reported independently in [9,10,12], in which FLDA were reported outperforming PCA significantly.

The drawback of FLDA is that it requires a large size of training samples for good generalization due to the singular problem of within-class scatter matrix. Usually, the image dimension can be reduced by using PCA before applying FLDA to image recognition. However, the discarded null space may contain significant discriminant information. Because of this, Tian et al. [13] proposed to use generalized inverse to solve the singular problem of within-class scatter matrix. Hong et al. [14] added a small perturbation to the within-class scatter matrix to ensure the non-singularity of within-class scatter matrix. This idea is intuitive but it is not clear how to set the best perturbation to the within-class scatter matrix. Chen et al. [15] proved that the null space of the within-class scatter matrix contains the most discriminative information when a small sample size problem takes place. Yu et al. [16] and Lu et al. [17] proposed, respectively, a direct FLDA (DFLDA) approach and a null space method to solve this problem. However, all the above methods are complex to calculate the projection basis vectors due to the image matrix to vector transformation. Recently, Ye et al. [18] proposed a two-dimensional FLDA (2D-FLDA) scheme for feature extraction. 2D-FLDA evaluates directly the within-class scatter matrix from the image matrix without image to vector transformation, and hence dilutes the singular problem of within-class scatter matrix.

Feature extraction using FLDA and PCA depends heavily on the estimation of the scatter matrices of training samples. In many applications, such as law enforcement, drive license, passport identification, each person has only one training sample. Under this condition, the within-class scatter matrix is zero and the FLDA-based algorithms fail, while the PCA-based algorithms have low recognition accuracy. Due to its significance to real-world applications, this challenging problem has rapidly emerged as an active research sub-area in face recognition, and many ad hoc techniques have been developed [19–26]. Wang et al. [19] presented a method to solve this problem by incorporating prior information of the within-class scatter from other subjects based on the assumption that human being exhibits similar intra-class variation. For that purpose, a generic training set with multiple samples per person is collected. Wu et al. [20] developed a $(PC)^2A$ approach. It linearly combines the first-order projection images and the original images to form new images, to which the Eigenface method is applied. Xie et al. [21] proposed to use Gabor filtered images and doubly nonlinear mapping to construct new face images and then use kernel PCA for training. Chen et al. [22] used the first- and second-order projection images and the original images to estimate sample covariance matrix. Zhanget et al. [23] proposed to perturb singular values of the images to augment training samples. These methods have better recognition accuracy than the standard Eigenface algorithm under single training image per person. However, the recognition accuracy is not good when FLDA is performed on the new training images. For FLDA, Chen et al. [24] divided the whole pattern into a set of non-overlapping sub-patterns to construct a new training sub-set and then evaluated the within-class scatter matrices and between-class matrices from these sub-patterns. Since this method needs to divide the test image into a set of non-overlapping sub-patterns, it will cost much time in classification.

In this paper, we present a singular value decomposition (SVD) based FLDA approach to solve the single training sample per person problem. We decompose the image into several basis images by using SVD. Most of the energy of an image concentrates on the several basis images associated with the large singular values and those basis images contribute a lot to the general appearance of the image. The other low-energy basis images, therefore, are the difference between the original image and the high-energy basis images. They are the edge-like components of the image and can reflect the variations of the same class images because the within-class variances emphasize more on the edge areas of the face images. Based on this observation, we partition the SVD of the image into two parts, the smooth general appearance part and the difference part. The later is used to approximately evaluate the within-class scatter matrix and the former is included in the calculation of between-class scatter matrix. Then 2D-FLDA is used to extract the discriminant features for recognition.

The rest of this paper is organized as follows. Section 2 briefly reviews the procedure of 2D-FLDA. Section 3 describes the proposed algorithm in detail. Section 4 performs extensive experiments on the well-known ORL, Yale, UMIST and FERET face databases. Section 5 concludes the paper.

2. Two-dimensional FLDA

FLDA is a popular feature extraction and discriminant approach to face recognition. It aims to find a set of projection vectors that separate the different classes as far as possible while compressing the same class as compact as possible. However, it is difficult to calculate accurately the generalized eigenvalues of within-class and between-class scatter matrices. Ye et al. [19] proposed 2D-FLDA, which directly estimates the scatter matrices from 2D images and uses the fisher discriminant criterion to compute the optimal projection vectors.

The 2D-FLDA can be briefly formulated as follows. Given K training samples $A_k \in R^{m\times n}$ ($k = 1,2,\ldots,K$) from C classes. The ith class C_i includes K_i samples and $\sum_{i=1}^{C} K_i = K$. Denote by $\overline{A}$ the mean image of all samples and by $\overline{A}_i$ the mean image of the ith class C_i. 2D-FLDA attempts to seek for a set of optimal discriminating vectors w_j ($j = 1,\ldots,d$) to construct a transform matrix $W = [w_1 w_2,\ldots w_d]$ by maximizing the following criterion

$$J(W) = \frac{tr\left(W^{\mathrm{T}} S_b W\right)}{tr\left(W^{\mathrm{T}} S_w W\right)}, \tag{1}$$

where superscript "T" denotes matrix transpose, tr denotes the trace of a matrix, S_b is the between-class matrix

$$S_b = \sum_{i=1}^{C} \frac{K_i}{K} \left(\overline{A}_i - \overline{A}\right)^{\mathrm{T}} (\overline{A}_i - \overline{A}) \tag{2}$$

and S_w is the within-class scatter matrix

$$S_w = \frac{1}{C} \sum_{k=1}^{K} \sum_{A_k \in C_i} \frac{1}{K_i} (A_k - \overline{A}_i)^{\mathrm{T}} (A_k - \overline{A}_i). \tag{3}$$

The maximization of Eq. (1) is equivalent to solve the generalized eigenvalue problem: $S_b W = \Lambda S_w W$, where Λ is a diagonal matrix with eigenvalues on the main diagonal.

After obtaining W, we can extract d discriminant features for any input image by projecting it onto W. As the number of discriminating vectors, d is at most $\min(C-1, n)$. Though FLDA has been widely used in face recognition, it cannot be used when each class has only one training image because this leads to a zero-matrix of S_w in the denominator of Eq. (1). In the following section, we will propose an approach to address this issue.

3. SVD-based FLDA with single training image per person

As well-known, it is the within-class scatter matrix S_w that makes the FLDA unstable when each class has only one training sample. So it is a key problem to approximately evaluate S_w for each class in using FLDA.

In this section, we propose a SVD-based solution to this problem. By using SVD, we decompose the image into two complementary parts: the first part is constructed by the SVD basis images associated with the several largest singular values, and the second part is constructed by the other low-energy basis images. The first part preserves most of the energy of an image and reflects the general appearance of the image. The second part is the difference between the original image and the first part and it can reflect, to some extent, the variations of the same class images. We use the second part to evaluate the within-class scatter matrix, while including the first part in the calculation of between-class scatter matrix. 2D-FLDA can then be used in the case of single training image per-class.

3.1. Image decomposition by using SVD

Given a face image $A \in R^{m \times n}$ and suppose $m \geqslant n$, we have the following expression according to SVD [27]:

$$A = \sum_{i=1}^{n} \sigma_i \cdot u_i \cdot v_i^{\mathrm{T}}, \tag{4}$$

where u_i and v_i are the ith column of $U \in R^{m \times m}$ and $V \in R^{n \times n}$, respectively. U and V are composed of the eigenvectors of AA^{T} and $A^{\mathrm{T}}A$, respectively. σ_i is the singular values of image A and we let $\sigma_i \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n$.

Eq. (4) can be explained as that A is composed of a set of basis images $F_i = \sigma_i \cdot u_i \cdot v_i^{\mathrm{T}}$, $i = 1, \ldots, n$. Most of the energy of A concentrates on the several basis images F_i with large singular values σ_i. Fig. 1 shows the percentage of energy of each basis image to the total image energy by averaging 100 face images. We can see that most of the energy of σ_i can be preserved by using 3–5 the most significant basis images F_i.

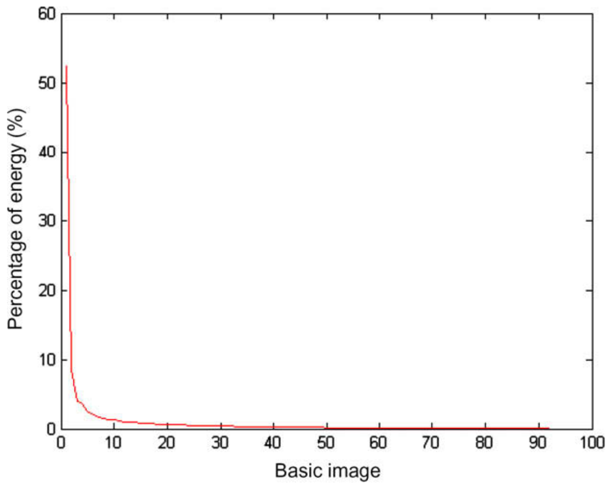

Fig. 1. The energy distribution of SVD basis images.

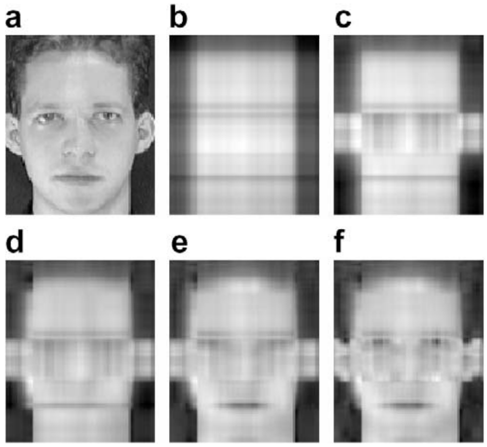

Fig. 2. (a) Original image; (b)–(f) are the constructed images by using 1–5 most significant SVD basis images.

Fig. 2 shows the original image and the reconstructed images by using 1–5 the most significant SVD basis images, i.e. $\widehat{A} = \sum_{i=1}^{k} \sigma_i \cdot u_i \cdot v_i^{\mathrm{T}}$, k = 1, 2,...,5, respectively. It is readily seen that the reconstructed images approach to the original image with increasing the number of basis images. If the number is bigger than 4, the difference between the reconstructed image and the original image is very small. As we can see in Fig. 2, the most significant SVD basis images reflect the general appearance of the face image.

Fig. 3 shows the original image and the difference image $\xi = A - \widehat{A}$, where $\widehat{A} = \sum_{i=1}^{3} \sigma_i \cdot u_i \cdot v_i^{\mathrm{T}}$. In other words, ξ is constructed by using the remaining non-significant SVD basis images: $\xi = \sum_{i=4}^{n} \sigma_i \cdot u_i \cdot v_i^{\mathrm{T}}$. It can be observed that image ξ reflects the edges, where the variations most probably occur, in the face image. In Fig. 4, we illustrate two samples of the same subject and their difference image. It is seen that the variations, as we expected, occur at the positions around the edges. This gives rise to an idea: we can use the image ξ to approximately evaluate the within-class scatter matrix, which cannot be directly calculated in the case of single training sample per-class.

3.2. Feature extraction and classification

Based on the observations and discussions in Section 3.1, from the original dataset that has only one training sample for each class, we can construct a new training set which includes two sample images for each class: one is the original image A and another one is the constructed image $\widehat{A} = \sum_{i=1}^{3} \sigma_i \cdot u_i \cdot v_i^{\mathrm{T}}$ by using the three most significant SVD basis images. We use A and $\widehat{A}$ to evaluate the between-class scatter matrix, and use $\xi = A - \widehat{A}$ to calculate the within-class scatter matrix. Then the Fisher discriminant criterion can be used for face identification.

Given C classes and that each class has only one sample $A_k \in R^{m \times n}$ (k = 1,...,C). Suppose $m \geqslant n$ and denote by $U_k \in R^{m \times m}$ and $V_k \in R^{n \times n}$ the matrices that are composed of the eigenvectors of $A_k A_k^{\mathrm{T}}$ and $A_k^{\mathrm{T}} A_k$, respectively. The constructed image $\widehat{A}_k$ by using the three most significant SVD basis images is

$$\widehat{A}_k = \sum_{i=1}^{3} \sigma_i^k \cdot u_i^k \cdot (v_i^k)^{\mathrm{T}}, \tag{5}$$

where u_j^k and v_j^k are the jth column of U_k and V_k, respectively, σ_i^k is the ith singular value of image matrix A_k and we let $\sigma_1^k \geqslant \sigma_2^k \geqslant \cdots \geqslant \sigma_n^k$.

After obtaining $\widehat{A}_k$, we have a new training set $\Theta = \left\{A_k \in R^{m \times n}, \widehat{A}_k \in R^{m \times n} | k = 1, 2, \ldots, C\right\}$. With Θ, it is readily to calculate the within-class scatter matrix S_w and between-class scatter matrix S_b in 2D-FLDA as

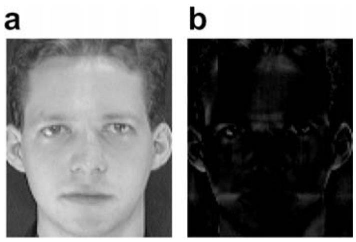

Fig. 3. (a) Original image; (b) the constructed images by using the non-significant SVD basis images.

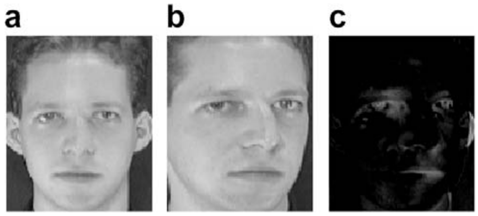

Fig. 4. (a) and (b) are two original images and (c) is the their difference.

$$S_w = \frac{1}{C}\sum_{k=1}^{C}\left[\left(A_k - \overline{A}_k\right)^{\mathrm{T}}\left(A_k - \overline{A}_k\right) + \left(\widehat{A}_k - \overline{A}_k\right)^{\mathrm{T}}\left(\widehat{A}_k - \overline{A}_k\right)\right], \tag{6}$$

$$S_b = \frac{1}{C}\sum_{k=1}^{C}\left(\overline{A}_k - \overline{A}\right)^{\mathrm{T}}\left(\overline{A}_k - \overline{A}\right), \tag{7}$$

where $\overline{A}$ denotes the global mean image and $\overline{A}_k$ denotes the mean of the kth class

$$\begin{cases} \overline{A}_k = \frac{1}{2}\left(A_k + \widehat{A}_k\right) \\ \overline{A} = \frac{1}{2C}\sum\limits_{k=1}^{C}\left(A_k + \widehat{A}_k\right) \end{cases}. \tag{8}$$

With (8), we can easily get that (6) is actually

$$S_w = \frac{1}{2C}\sum_{k=1}^{C}\xi_k^{\mathrm{T}}\xi_k, \tag{9}$$

where $\xi_k = A_k - \widehat{A}_k$.

The optimal projection matrix W on dataset Θ is composed of the eigenvectors of $(S_w)^{-1} S_b$. Having obtained W, we can extract the discriminant features Z_k from A_k by

$$Z_k = A_k \times W, \quad k = 1, \ldots, C. \tag{10}$$

Given a test image A^*, we first project it onto W to get the discriminant feature Z^*, then select the nearest neighbor classifier to identify A^*. Here, the distance between Z^* and Z_k is

$$D(Z^*, Z_k) = \sum_{l=1}^{d}\left\|z_l^* - z_l^k\right\|_2, \quad k = 1, \ldots, C, \tag{11}$$

where z_l^* and z_l^k are the lth column of Z^* and Z_k, respectively. If

$$D(Z^*, Z_p) = \arg\left\{\min_k D(Z^*, Z_k)\right\}, \tag{12}$$

then the test face image A^* is identified belonging to the pth class.

4. Experimental results

In the experiments, we evaluate the performance of the proposed method using four well-known face databases: Yale [28], FERET [29], ORL [30] and UMIST [31]. The Yale database is used to evaluate the performance of the proposed method under expression and illumination variations; the FERET database is used to test the performance under expression variations; the ORL database is used to test the performance under size and rotation variations; and the UMIST database is selected to evaluate the performance under pose variations.

The Yale database contains images from 15 individuals, each providing 11 different samples. The images have variations in lighting conditions (left-light, center-light, right-light), facial expressions (normal, happy, sad, sleepy, surprised, and wink), and occlusion (with/without glasses). In this database, preprocessing to locate the faces was applied. Original images

Fig. 5. Some images for one subject in the Yale database.

were normalized (in scale and orientation) such that the two eyes were aligned at the same position. The facial area was then cropped for matching. The size of each cropped image is 32×32. Fig. 5 shows some cropped images of one subject.

The FERET face database consists of 3280 gray-level frontal view face images from 1010 persons. In the experiment, we selected 400 images from 200 persons for testing. Each person has two images (**fa** and **fb**) which are obtained at different times and with different facial expressions. All the images are manually cropped to the size of 40×40.

The ORL database contains samples from 40 individuals, each providing 10 different images. For some subjects, the images were taken at different times. The facial expressions (open or closed eyes, smiling or non-smiling) and occlusion (glasses or no glasses) also vary. The images were taken with a tolerance for tilting and rotation up to 20°. There is also some variation in the scale of up to 10%. All images are grayscale and normalized to a resolution of 112×92 pixels. Fig. 6 shows five sample images of one person in the ORL database.

The UMIST database was built by the University of Manchester Institute of Science and Technology. It is a multi-view database, consisting of 575 images from 20 people, each covering a wide range of poses from profile to frontal views. In the experiment, 19 poses per person were selected to form a new sub-database of 380 images. The image size is 112×92. Fig. 7 shows some images of one person.

In the experiments on Yale, ORL and UMIST databases, the first image of each subject was selected for training with the remaining images for testing. In the FERET **fafb** database, we used **fa** images for training and **fb** images for testing. Three methods, which were proposed to solve the single training sample per person problem, are used for comparison. The first method is the projection approach [22]; the second is the singular value perturbation approach [23]; and the third is to divide an image into several non-overlap blocks [24]. In all the four schemes, 2D-FLDA is used to extract the discriminant features and the nearest classifier is used to classify the input image. All the parameters in the three comparison algorithms are set as the same as those in [22,23] and [24], respectively.

Fig. 8 plots the recognition accuracies by different schemes under the different number of projection vectors on the four databases. Since [24] is a block based scheme, its associated number of projection vectors is usually small. Table 1 lists the top recognition accuracy of the four methods. On Yale database, the top recognition accuracy is 18.67% for the first method, 23.33% for the second method, 32.00% for the third method, and 34.67% for the proposed method. On **fafb** database, the top recognition accuracies are 83.00%, 86.50%, 89.50% and 90.50% for the four methods, respectively. On ORL database, the top recognition accuracies are 44.17%, 46.39%, 70.83% and 75.556% for the four methods, respectively. On UMIST database, the top recognition accuracies are 18.61%, 52.5%, 57.22%, and 61.39%, respectively. It can be seen that the proposed method achieves higher recognition accuracy than the other three methods. Comparing with [22] and [23], the proposed method improves more than 10% in recognition accuracy except for the FERET **fafb** database. This is because these two methods were designed for improving recognition accuracy of PCA-based algorithms. The proposed method improves about 5% in recognition accuracy than the method in [24], which was designed for FLDA.

4.1. Discussions

In the previous development of the proposed scheme, SVD is used to evaluate the difference image ξ. Actually, other tools, such as discrete cosine transform (DCT) and wavelet transform, can also be used for this purpose. Here we present some experimental results by using DCT as the image decomposition tool in the proposed method. In experiments, we threshold the DCT coefficients to construct the smooth general apparent part $\widehat{A}$ and the difference image is $\xi = A - \widehat{A}$. The threshold is set as 2, 3, 2 and 2 for the Yale, FERET, ORL and UMIST databases, respectively. Table 2 lists the top recognition accuracies on

Fig. 6. Some images of one subject in the ORL database.

Fig. 7. Some images of one subject in the UMIST database.

the four databases with the same training samples and testing images as Section 4.1. It can be seen that by using DCT as the image decomposition tool, the recognition results are very similar to those by using SVD.

The proposed idea can also be used in LPP [32] when each object has one training sample. By using SVD, we divide the image into two parts $\widehat{A}$ and ξ and then apply LPP to them. Usually, PCA is first used to reduce dimensionality of image space

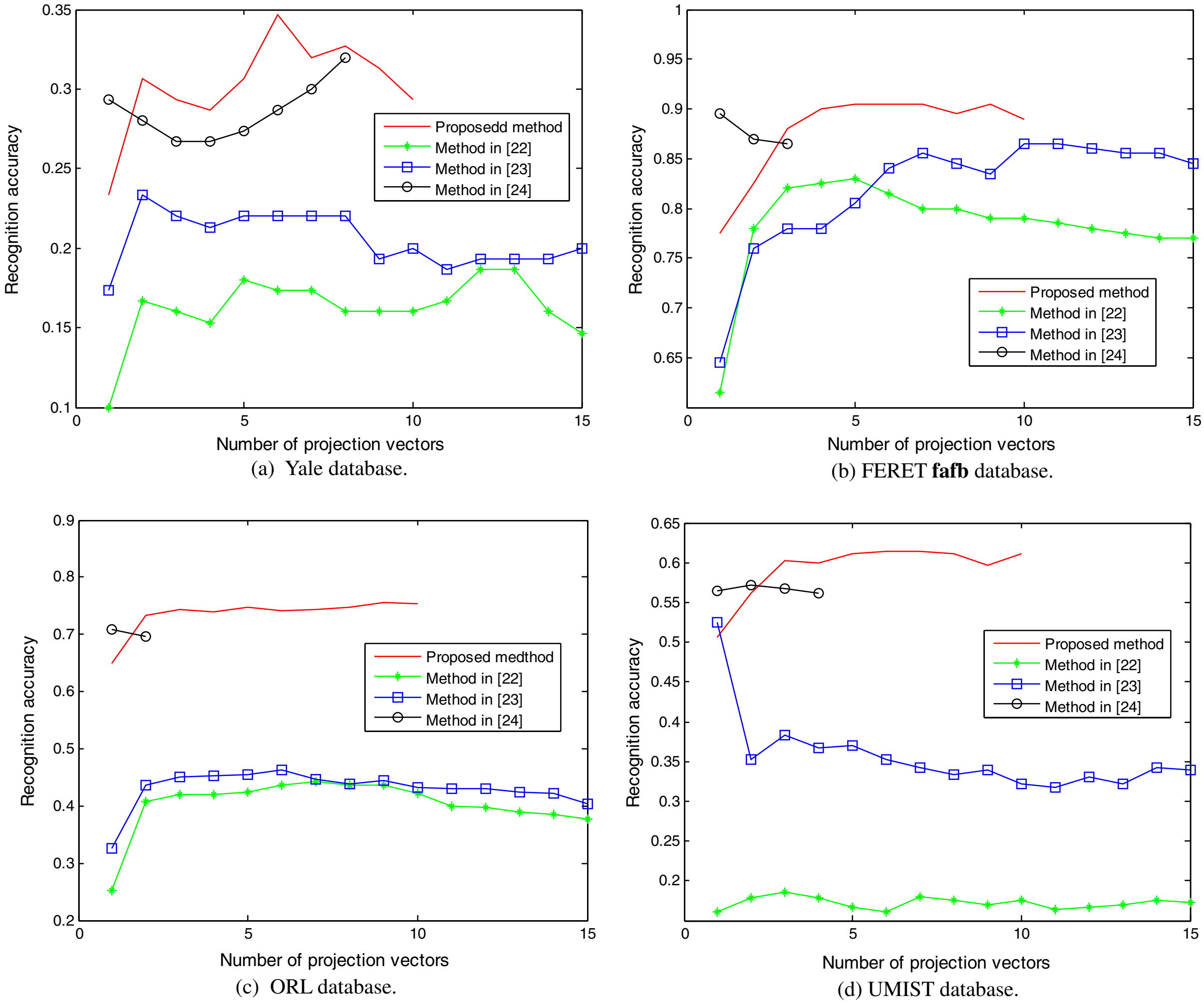

Fig. 8. Curves of recognition accuracies by different methods. (a) Yale database; (b) **fafb** database; (c) ORL database; (d) UMIST database.

Table 1
The top recognition accuracies of the four methods on the four databases

Database	Method in [22] (%)	Method in [23] (%)	Method in [24] (%)	Proposed method (%)
Yale	18.67	23.33	32.00	34.67
FERET fafb	83.00	86.50	89.50	90.50
ORL	44.17	46.39	70.83	75.56
UMIST	18.61	52.50	57.22	61.39

Table 2
The top recognition accuracies of the proposed method by using SVD and DCT

Database	Yale (%)	FERET fafb (%)	ORL (%)	UMIST (%)
SVD	34.67	90.50	75.56	61.39
DCT	36.67	91.00	76.39	59.44

Table 3
The top recognition accuracies by applying the proposed method to LPP

PCA ration	0.85		0.9		0.95	
	LPP (%)	LPP + SVD (%)	LPP (%)	LPP + SVD (%)	LPP	LPP + SVD (%)
Yale	24.00	30.00	30.67	32.67	31.33	35.33
FERET **fafb**	85.00	87.00	84.00	89.00	85.00	85.50
ORL	66.11	70.83	68.06	68.61	70.00	68.06
UMIST	51.11	60.28	60.56	55.28	54.72	65.56

before applying LPP. In experiments, we use PCA ration to decide how many energy has been reserved. Table 3 lists the top recognition accuracies of original LPP method and the proposed "SVD + LPP" method. The training and testing images are the same as those in Section 4.1. We see that in most cases the recognition accuracy of LPP can be improved by using the proposed scheme. However, it needs further investigation about how to better apply the proposed idea in this paper to LPP and other local embedding methods.

5. Conclusion

This paper presented a simple but efficient method to solve the problem of single training image per person when using Fisher linear discriminant analysis (FLDA). By using singular value decomposition (SVD), we divided the face image A into two complementary parts: the smooth general appearance part $\widehat{A}$ and the difference part $\xi = A - \widehat{A}$. The first part $\widehat{A}$ was used in the calculation of between-class scatter matrix and ξ was used to evaluate the within-class scatter matrix, which cannot be directly calculated in the case of single training sample per-class. Experimental results verify that the proposed method is not only feasible but also achieves better recognition performance.

Acknowledgement

We thank the anonymous reviews for their constructive comments and suggestions. We thank Dr. He Xiaofei for providing the cropped images of Yale database and thank Dr. Yang Jian for providing the cropped images of FERET. This project was supported by the 111 Project of China (B08038).

References

[1] X. Wang, X.O. Tang, A unified framework for subspace face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (9) (2004) 1222–1228.
[2] W. Zhao, R. Chellappa, A. Rosenfeld, P. Phillips, Face recognition: a literature survey, ACM Comput. Surveys 25 (2003) 399–458.
[3] S.C. Yan, D. Xu, B. Zhang, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 40–51.
[4] M. Kirby, L. Sirovich, Application of the KL procedure for the characterization of human faces, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1) (1990) 103–108.
[5] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1) (1991) 72–86.
[6] K. Sung, T. Poggio, Example-based learning for view-based human face detection, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1) (1998) 39–51.
[7] J. Yang, D. Zhang, A. Frangi, J.Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004) 131–137.
[8] S. Fidler, D. Skoaj, A. Leonardis, Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling, IEEE Trans. Pattern Anal. Mach. Intell. 28 (3) (2006) 337–350.
[9] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class special linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[10] D. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Trans. Pattern Anal. Machine Intell. 18 (1996) 831–836.
[11] K. Etemad, R. Chellappa, Discriminant analysis for recognition of human face images, J. Opt. Soc. Am. A 14 (1997) 1724–1733.
[12] W. Zheng, J. Lai, P. Yuen, GA-Fisher: a new LDA-based face recognition algorithm with selection of principal components, IEEE Trans. Sys. Man Cyb. Part B. 35 (5) (2005) 1065–1078.
[13] Q. Tian, Comparison of statistical pattern-recognition algorithm for hybrid processing, II. Eigenvector-based algorithm, J. Opt. Sot. Am. A.5 (1988) 1670–1672.
[14] Z.Q. Hong, J.Y. Yang, Optimal discriminant plane for a small number of samples and design method of classifier on the plane, Patter Recognit. 24 (4) (1991) 317–324.
[15] L. Chen, H. Liao, M. Ko, J. Lin, G. Yu, A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognit. 33 (10) (2000) 1713–1726.
[16] H. Yu, J. Yang, A direct LDA algorithm for high-dimensional data with application to face recognition, Pattern Recognit. 34 (2001) 2067–2070.
[17] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Face recognition using LDA-based algorithms, IEEE Trans. Neural Networks 14 (1) (2003) 195–200.
[18] J. Ye, R. Janardan, Q. Li, two-dimensional linear discriminant analysis, in: Proceedings of Neural Information Processing Systems (NIPS), Vancouver, BC, Canada, 2004.
[19] J. Wang, K.N. Plataniotis, A.N. Venetsanopoulos, Selecting discriminant eigenfaces for face recognition, Pattern Recognit. Lett. 26 (10) (2005) 1470–1482.
[20] J. Wu, Z. Zhou, Face recognition with one training image per person, Pattern Recognit. Lett. 23 (2002) 1711–1719.
[21] X. Xie, K. Lam, Gabor-based kernel PCA with doubly nonlinear mapping for face recognition with a single face image, IEEE Trans. Image Processing 15 (9) (2006) 2481–2492.
[22] S.C. Chen, D.Q. Zhang, Z.H. Zhou, Enhanced $(PC)^2A$ for face recognition with One training image per person, Pattern Recognit. Lett. 25 (10) (2004) 1173–1181.

[23] D. Zhang, S.C. Chen, Z.H. Zhou, A new face recognition method based on SVD perturbation for single example image per person, Appl. Math. Comput. 163 (2) (2005) 895–907.
[24] S. Chen, J. Liu, Z.H. Zhou, Making FLDA applicable to face recognition with one sample per person, Pattern Recognit. 37 (1) (2004) 553–1555.
[25] X. Tan, S.C. Chen, Z.H. Zhou, F. Zhang, Face recognition from a single image per person: a survey, Pattern Recognit. 39 (2006) 1725–1745.
[26] Z.Q. Zhao, D.S. Huang, B.Y. Sun, Human face recognition based on multiple features using neural networks committee, Pattern Recognit. Lett. 25 (12) (2004) 1351–1358.
[27] G. Golub, C. Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, MD, 1983.
[28] Yale Univ. Face Database <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>, 2002.
[29] P.J. Phillips, H. Moon, P. Rauss, The FERET evaluation methodology for face recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1103.
[30] ORL database <http://www.cam-orl.co.uk/facedatabase.html>, 2002.
[31] UMIST Face Database, Daniel Graham <http://images.ee.umist.ac.uk/danny/database.html>, 2005.
[32] X.F. He, S.C. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 1–13.