Digital Signal Processing 22 (2012) 140–146

Contents lists available at SciVerse ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Face recognition using difference vector plus KPCA

Ying Wen [a,b,*], Lianghua He [b,c], Pengfei Shi [d]

[a] Department of Computer Science and Technology, East China Normal University, Shanghai, 200062, China
[b] Pediatric Brain Imaging Laboratory, Columbia University in the City of New York, NY 10032, USA
[c] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai, 200092, China
[d] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, 200030, China

ARTICLE INFO

Article history:
Available online 6 September 2011

Keywords:
Face recognition
Common vector
Kernel PCA

ABSTRACT

In this paper, a novel approach for face recognition based on the difference vector plus kernel PCA is proposed. Difference vector is the difference between the original image and the common vector which is obtained by the images processed by the Gram–Schmidt orthogonalization and represents the common invariant properties of the class. The optimal feature vectors are obtained by KPCA procedure for the difference vectors. Recognition result is derived from finding the minimum distance between the test difference feature vectors and the training difference feature vectors. To test and evaluate the proposed approach performance, a series of experiments are performed on four face databases: ORL, Yale, FERET and AR face databases and the experimental results show that the proposed method is encouraging.

1. Introduction

Face recognition by machine has been started since 1970s, due to military, commercial, and law enforcement applications, currently becoming an active and important research area. Face images are very sensitive to the variations such as face appearance, pose and expression variations, which impact on the recognition result [1–3]. Moreover, the higher the dimension of face image, the more the time-consumption of recognition. These issues make face recognition a difficult task.

Feature extraction acts as a vital role for face recognition. Principal component analysis (PCA) is a well-known method for feature extraction [4–6]. By calculating the eigenvectors of the covariance matrix of the original inputs, PCA linearly transforms a high-dimensional input vector into a low-dimensional one whose components are uncorrelated [7,8]. Although PCA has many advantages, it also has many shortcomings, such as its sensitiveness to noise, and its limitation to data description. To eliminate these shortcomings, many methods have been proposed to improve PCA algorithm. These methods mainly focus on two aspects. On the one hand, the sensitiveness of the traditional PCA to noise seriously affects its precision. Many scholars analyzed the robustness of PCA in detail, and presented many improved PCA algorithms [9–11]. On the other hand, for the traditional PCA, principal components are determined exclusively by the second-order statistics of the input data, which can only describe the data with smooth Gaussian distribution. In order to describe the data with non-Gaussian distribution, many researchers such as Karhunen et al. introduced an appropriate nonlinear processes into the traditional PCA and developed a nonlinear PCA [12] according to the distribution of the input samples. In these improved PCA algorithms, kernel based PCA (KPCA) [13] proposed by Scholkopf et al. is a state-of-the-art one as a nonlinear PCA algorithm. KPCA utilizes kernel function to gain the random high-order correlation between input variants, and finds the principal components needed through the inner production between input data. KPCA not only can successfully describe the data with Gaussian distribution, but also can commendably describe the data with non-Gaussian distribution. More and more people are interested in this field and have carried out some relevant researches [14–16].

Recently, common vector was proposed and originally introduced for isolated word recognition problems in the case where the number of samples in each class was less than or equal to the dimensionality of the sample space [17,18]. Common vector presents the common properties of a training set. Inspired by this idea, Hakan Cevikalp et al. [19] proposed an approach of discriminative common vectors (DCV) for face recognition, which uses the subspace methods and the Gram–Schmidt orthogonalization procedure to obtain the discriminative common vectors. Afterwards, Yunhui He et al. [20] also presented an algorithm called kernel DCV by combining DCV with kernel method (KDCV). In addition, they also proposed a common face approach for face recognition (CVP) [21]. The recognition result of CVP was derived by finding the minimum distance between a face image and the common vectors. And then, they also extend kernel method to CVP (KCVP).

In this paper, we propose a new face recognition method based on difference vectors plus KPCA approach. Common vector of each

* Corresponding author at: Pediatric Brain Imaging Laboratory, Columbia University in the City of New York, NY 10032, USA.
E-mail address: yw2365@columbia.edu (Y. Wen).

1051-2004/$ – see front matter
doi:10.1016/j.dsp.2011.08.004

class is gotten by the procedure of the Gram–Schmidt orthogonalization, while difference vector is derived by the difference between the original image and the common vector. The optimal feature vectors are obtained by KPCA procedure of difference vectors. Recognition result is derived from finding the minimum distance of the feature vectors. Experiments on ORL, Yale, FERET and AR face databases indicate that the proposed approach is encouraging.

The rest of this paper is organized as follows: Section 2 mainly presents the algorithm of the proposed approach; in Section 3, some experimental results on some famous face databases are presented; finally, conclusion is also given.

2. Algorithm description

2.1. Common vector

Assume that there are C classes, and each class contains N samples. Let $x_j^i, i = 1, 2, \ldots, C, \ j = 1, 2, \ldots, N$ be an n-dimensional column vector which denotes the jth sample from the ith class. There are a total of $M = NC$ samples in the training set. Construct B^i of the ith class $i = 1, 2, \ldots, C$ whose columns span a difference subspace for the ith class is defined as follows:

$$B^i = \left[b_1^i, b_2^i, \ldots, b_{N-1}^i\right] \tag{1}$$

where $b_k^i = x_{k+1}^i - x_1^i$, $k = 1, 2, \ldots, N-1$, x_1^i is called reference vector which can be randomly selected from the ith class, and the first sample is selected in this paper.

By performing Gram–Schmidt orthogonalization procedure, the orthonormal vector set $y_1^i, y_2^i, \ldots, y_{N-1}^i$ which spans the difference subspace $L(B^i)$ is obtained. Then a sample x_k^i randomly selected from the ith class is projected on the orthonormal vector y_k^i $(k = 1, 2, \ldots, N-1)$, and the summation of the projection is computed as follows:

$$s^i = \langle x_k^i, y_1^i \rangle y_1^i + \langle x_k^i, y_2^i \rangle y_2^i + \cdots + \langle x_k^i, y_{N-1}^i \rangle y_{N-1}^i \tag{2}$$

where $\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$ denotes the scalar product of $x = (x_1, x_2, \ldots, x_n) \in R^n$ and $y = (y_1, y_2, \ldots, y_n) \in R^n$. Then common vector x_{common}^i of the ith face class is derived as follows:

$$x_{common}^i = x_k^i - s^i \tag{3}$$

It was proved that common vector x_{common}^i is unique and independent of the randomly selected sample x_k^i (Appendix A). Therefore, common vector x_{common}^i represents the common invariant properties of the ith face class. As the above statement, we obtain C common vectors.

2.2. Principal component analysis

Principal component analysis is a popular technique for feature extraction and dimensionality reduction. Suppose that $X = \{x_p;\ \ p = 1, 2, \ldots, M\}$ is a set of centered observations of an n-dimensional zero-mean variable. Let

$$\sum_{p=1}^{M} x_p = 0 \tag{4}$$

The covariance matrix of the variable can be estimated as follows:

$$C_x = \frac{1}{M} \sum_{p=1}^{M} x_p x_p^T \tag{5}$$

PCA aims at making the covariance matrix C_x in Eq. (5) be diagonal. It leads to an eigenvalue problem:

$$\lambda u = C_x u \tag{6}$$

where λ is the eigenvalues of C_x and u is the corresponding eigenvectors.

PCA linearly transforms x into a new one z:

$$z = u^T x \tag{7}$$

The new components are called principal components. By using only the first several eigenvectors sorted in descending order of the eigenvalues, the number of principal components in z can be reduced. So PCA has the dimensional reduction characteristic and the principal components are uncorrelated.

2.3. Relation between common vector and PCA

The face image can be regarded as a summation of common vectors of the ith face class which represents the common invariant properties of the ith face class, while a difference vector represents the specific properties of a face image. However, what does the difference vector imply?

The eigenvalues of the ith class covariance matrix C_x are nonnegative and they can be written in decreasing order: $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. Let $u_1, u_2, \ldots, u_n$ be the orthonormal eigenvectors corresponding to these eigenvalues. The first $(m-1)$ eigenvectors of the covariance matrix correspond to the nonzero eigenvalues.

Let $Ker\, C_x$ be the space of all eigenvectors corresponding to the zero eigenvalues of C_x, and $B^\perp$ be the orthogonal complement of the difference subspace B. Further, since the space B is $(m-1)$-dimensional, the space $B^\perp$ is $(n-m+1)$-dimensional [25].

$$Ker\, C_x = \left\{x \in R^{n \times 1}\colon\ C_x x = 0\right\}$$
$$B^\perp = \left\{x \in R^{n \times 1}\colon\ \langle x, b \rangle = 0\ \forall b \in B\right\}$$

From Refs. [18,26], some conclusions can be obtained:

1. $Ker\, C_x = B^\perp$ (8)
2. Because x_{common} is orthogonal to any vector in the difference subspace. Let $(x_i - x_j) \in B$, so
$$\langle x_i - x_j, x_{common} \rangle = 0 \tag{9}$$
3. $C_x x_{common} = 0$ (10)
4. $B^\perp = span[u_m, u_{m+1}, \ldots, u_n]$ or
$$B = span[u_1, u_2, \ldots, u_{m-1}] \tag{11}$$

Based on the conclusions (1)–(4), any feature vector x can be written as

$$x = \langle x, u_1 \rangle u_1 + \cdots + \langle x, u_{m-1} \rangle u_{m-1} + \langle x, u_m \rangle u_m + \cdots + \langle x, u_n \rangle u_n \tag{12}$$

or

$$x = x^* + x^\perp \tag{13}$$

where

$$x^* = \langle x, u_1 \rangle u_1 + \cdots + \langle x, u_{m-1} \rangle u_{m-1}, \quad x^* \in B$$

and

$$x^\perp = x_{common} = \langle x, u_m \rangle u_m + \cdots + \langle x, u_n \rangle u_n, \quad x^\perp \in B^\perp$$

So, for any feature vector x, the common vector x_{common} can also be written as

$$x_{common} = x - x^* \tag{14}$$

As seen from the above derivations, common vector can be determined by using the eigenvectors corresponding to the zero or nonzero eigenvalues of the covariance matrix C_x. But it should be noted that common vector does not include information in the directions corresponding to the nonzero eigenvalues. Common vector is orthogonal to all the eigenvectors that correspond to the nonzero eigenvalues of the covariance matrix. Common vector is unique for its class and contains all the common invariant features of its own class.

For an ith class sample, difference vector is the difference between the sample and common vector of the ith class, which is defined by:

$$x_{diff}^i = x_{sample}^i - x_{common}^i \tag{15}$$

As we know, common vector that corresponds to the zero eigenvalues of the covariance matrix contains the invariant features of the class, while difference vector x_{diff}^i reserves all the components of a feature vector that are along the eigenvectors corresponding to the nonzero eigenvalues of the covariance matrix. So x_{diff}^i contains more details of the within-individual variations. In our method, difference vectors replacing the original face images are used for face recognition.

2.4. Kernel principal component analysis

KPCA is an approach to generalize linear PCA into nonlinear case using the kernel method. The idea of KPCA is to firstly map the original input vectors x_p into a high-dimensional feature space $\Phi(x_p)$ and then to calculate the linear PCA in $\Phi(x_p)$. By mapping x_p into $\Phi(x_p)$ whose dimension is assumed to be larger than the number of training samples M,

$$\Phi : x \to \Phi(x) \tag{16}$$

Assume that the mapped observations are centered, i.e.,

$$\sum_{p=1}^{M} \Phi(x_p) = 0 \tag{17}$$

Then the covariance matrix in the feature space is:

$$\Sigma = \frac{1}{M} \sum_{p=1}^{M} \Phi(x_p) \Phi(x_p)^T \tag{18}$$

The corresponding eigenvalue problem is:

$$\lambda \mu = \Sigma \mu \tag{19}$$

Eq. (19) can be transformed to the eigenvalue problem of kernel matrix:

$$\delta_p \alpha_p = K \alpha_p, \quad p = 1, \ldots, M \tag{20}$$

where K is the $M \times M$ kernel matrix, δ_p is one of the eigenvalue of K and α_p is the corresponding eigenvector of K. Finally, the principal components for x_p are calculated by [22]:

$$Y_p = \sum_{p=1}^{M} \frac{\alpha_p}{\sqrt{\delta_p}} K(x_p, x) \tag{21}$$

By the mapping Φ, we assume that an original nonlinear problem in the input space can be transformed to a linear problem in the high-dimensional feature space. However, it is impossible to compute the matrix K directly without carrying out the mapping Φ. Fortunately, for certain mapping Φ and corresponding feature spaces, there is a highly effective trick for computing the dot product $(\Phi(x) \cdot \Phi(y))$ in feature spaces using kernel functions. If the mapping Φ satisfies the Mercer's condition, the dot product can be replaced by a kernel function as follows:

$$K(x, y) = \big(\Phi(x) \cdot \Phi(y)\big) \tag{22}$$

which allows us to compute the value of the dot product in the high-dimensional feature space without having to carry out the mapping Φ explicitly. The following polynomial kernel is a commonly used kernel function:

$$K(x, y) = (x \cdot y)^l \tag{23}$$

where l is any positive integer.

In general, Gaussian kernel gives good performance when the optimal parameter is used. However, the optimal parameter selection is difficult. It is reported that normalized polynomial kernel gives the comparable performance with Gaussian kernel [23]. In addition, parameter dependency of normalized polynomial kernel is low. Therefore, we use normalized polynomial kernel as the kernel function [23,24]. Normalized polynomial kernel is defined as

$$K(x, y) = \frac{(1 + x^T y)^l}{\sqrt{(1 + x^T x)^l (1 + y^T y)^l}} \tag{24}$$

In the following experiments, parameter l is set to 3 by preliminary experiment.

2.5. Classification method

In our method, difference vector of each sample is taken as the input. The optimal feature vectors are obtained by KPCA procedure on difference vectors. For a training difference vector $A_j^i = x_j^i - x_{common}^i$, let

$$Z_{train}^i = \frac{1}{N} \sum_{j=0}^{N} Z_j^i = \frac{1}{N} \sum_{j=0}^{N} \sum_{p=1}^{d} \frac{\alpha_p}{\sqrt{\delta_p}} K\big(A_j^i, x_{diff}^p\big),$$
$$i = 1, 2, \ldots, C; \; j = 1, 2, \ldots, N; \; p = 1, 2, \ldots, d \tag{25}$$

where x_{diff}^p is a training sample difference vector and Z_j^i is a feature vector obtained from the jth difference vector of the ith class processed by KPCA. And Z_{train}^i is the average feature vector of the ith class.

For a test difference vector $A_{test}^i = x_{test} - x_{common}^i$, its feature vector is also gotten by:

$$Z_{test}^i = \sum_{p=1}^{d} \frac{\alpha_p}{\sqrt{\delta_p}} K\big(A_{test}^i, x_{diff}^p\big),$$
$$i = 1, 2, \ldots, C; \; p = 1, 2, \ldots, d \tag{26}$$

Suppose that there are C classes. So we get C average feature vectors Z_{train}^i of training samples and C feature vectors Z_{test}^i of a testing sample. The classifier is adopted as follows:

$$D^r(x) = min\big\{D^i(x)\big\} = min\big\{\big\| Z_{test}^i - Z_{train}^i \big\|^2\big\},$$
$$i = 1, 2, \ldots, C \tag{27}$$

If there is a minimum dissimilarity among C dissimilarities $D^i(x)$, it indicates that the class with the minimum dissimilarity is the recognition result. Then the testing sample belongs to the r class.

In the paper, a method for face recognition based on difference vector plus KPCA is proposed (we call the method DV-KPCA here after). If PCA is taken as feature extraction technique, the method is defined as DV-PCA. DV-KPCA algorithm performs the following computation:

Fig. 1. Face images on ORL database (row 1). Common vectors (row 2).

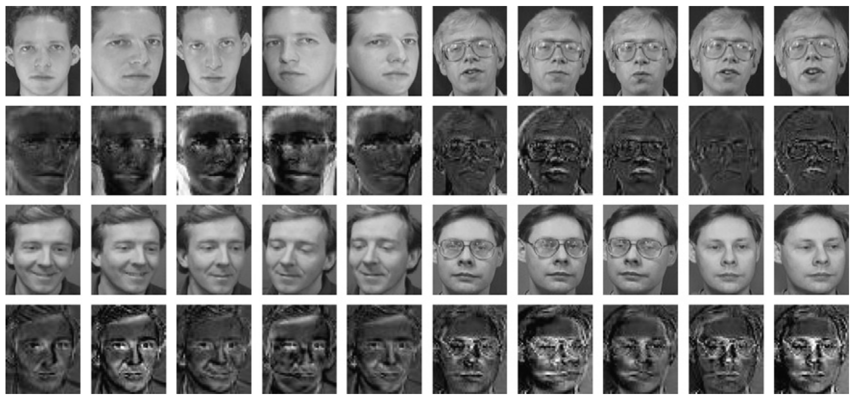

Fig. 2. Face images of four classes and the respective difference vectors.

Step 1: Construct difference subspace B^i of class i ($i = 1, 2, \dots, C$). Apply the Gram–Schmidt orthogonalization to obtain an orthonormal basis for B^i. Choose a sample from each class and project it onto the orthogonal complement of B^i to obtain common vector for each class. By these procedures, we get C common vectors and obtain all training and testing difference vectors.

Step 2: Construct difference vector kernel matrix and center the kernel matrix K. Computer the nonzero eigenvalues and corresponding eigenvectors of K and the coefficients in terms of the corresponding eigenvector and eigenvalue of kernel matrix. Use these eigenvectors to form the projection vectors which are used to obtain feature vectors.

Step 3: Implement the recognition work according to the classification criteria.

3. Experiments and results

In this section, we experimentally evaluate the recognition performance of our proposed approach on four well-known face databases: ORL, Yale, FERET and AR. All experiments are carried out on a PC machine with P4 2.8 GHz CPU and 512 MB RAM memory under Matlab 7 platform.

3.1. Experiments on ORL face database

In this group of experiments, the Olivetti–Oracle Research Lab (ORL) face database is adopted (http://www.uk.research.att.com/facedatabase.html) and used to test the performance of face recognition algorithms under the condition of minor variation of scaling and rotation. ORL face database contains 112×92 sized 400 frontal faces: 10 tightly, cropped images of 40 individuals with variation in pose, illumination, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses).

We randomly select five face images from each face class. Fig. 1 shows the original face images on ORL database (row 1) and common vector of each face class (row 2). The experiments verify that common vector possesses the common invariant properties of the class. It also indicates that common vector is independent of the randomly selected sample. Fig. 2 shows that four classes face images and their corresponding difference vectors. It can be seen from Fig. 2 that the difference face images represent the specific properties of the respective face images due to the variants such as face appearance, pose and expression variants. In this paper, we use difference face images for face recognition.

Table 1
Performance comparison of four methods on ORL database.

Method	Average recognition rate (%)	Dimension	Running time (ms)
PCA	93.32	80	21.78
LDA	94.55	39	10.05
DV-PCA	95.55	100	26.84
DV-KPCA	97.05	80	22.32

In the following experiments, the normalized polynomial kernel of DV-KPCA is adopted and the parameter l is equal to 3. We employ two groups to test the recognition performances of our proposed approach. First, we experimentally evaluate our proposed approach with PCA and LDA approaches which employ the technique for dimensionality reduction. And then, we experimentally evaluate our proposed approach with CVP, KCVP, DCV, KDCV approaches which employ the common vector approach.

3.1.1. Experiments of the approaches with lower-dimension technique

In the first experiment, we randomly select five images of each class as the training set and the rest as the testing set. This experiment is repeated 20 times. The average recognition rates of different algorithms are, respectively, summarized in Table 1. It also gives the comparisons of four approaches on the recognition rate, corresponding dimension of feature vector and running time. We see from Table 1 that the recognition rate of DV-KPCA is better than that of PCA, LDA, and DV-PCA, but not in terms of the running time.

Now, let us design a series of experiments to compare PCA, LDA, DV-PCA, and DV-KPCA methods under conditions where the sample size is varied. Here, four tests are performed with a varying number of training samples. we randomly select some images from each class to construct the training set and the remaining images as the testing set. To ensure the sufficient training samples, at least 4 samples are used. All of experiments are repeated 20 times. Table 2 shows the comparison of the average error rates of different

Table 2
Average error rate (%) comparison of different approaches on ORL database.

Method/number of training samples	4	5	6	7
PCA	8.05	5.68	4.78	3.25
LDA	7.23	5.45	4.03	3.04
DV-PCA	7.34	4.45	3.97	2.21
DV-KPCA	6.55	2.95	2.78	2.08

Table 3
Average recognition rate (%) comparison of six methods on ORL database.

Method/number of training samples	3	5	7
CVP [13]	89.21	93.6	96.67
KCVP [13]	90.39	95.50	97.75
DCV [11]	90.70	96.11	97.77
KDCV [11]	90.54	96.32	97.65
DV-PCA	88.45	95.55	97.79
DV-KPCA	89.15	97.05	97.92

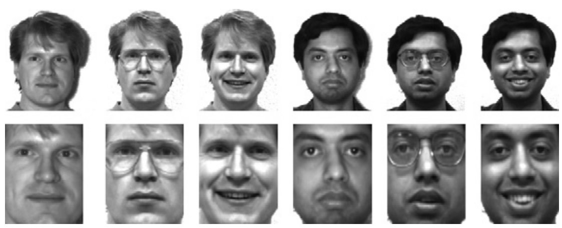

Fig. 3. First row: some original images on Yale database. Second row: normalized images.

approaches on ORL database for varying number of training samples. The error rate of the proposed approach is the minimum in all tests. As the number of training samples increases, the error rate decreases obviously.

3.1.2. Experiments of the approaches with common vector

We design a series of experiments to compare the performance of the proposed method, CVP, KCVP, DCV, and KDCV methods under conditions where the training sample size is varied. We randomly select $3, 5, 7$ samples from each class for the training, and the remaining samples of each class for the testing. The experiments are also repeated 20 times. The averaging recognition rates of these experiments are listed in Table 3. From the table, we also see that the recognition rate of DV-KPCA is comparable to that of CVP, KCVP, DCV, KDCV, and DV-PCA. Actually, from the experimental results, the recognition performances of DCV and DV-KPCA are comparable and similar. In this paper, we focus on the verification which is the difference vector good for face recognition, while it is undeniable that DCV is also indeed a good method. Although different feature extraction, the good performances of DCV and DV-KPCA are really obtained.

3.2. Experiments on Yale database

Yale database (http://cvc.yale.edu/projects/yalefaces/yalefaces.html) contains 165 grayscale images of 15 individuals, each of which is cropped with the size of 100×80. There are 11 images per subject, one per different facial expression or lighting configurations. We considered this database in order to evaluate the performance of methods under the condition when facial expression and lighting conditions are changed. Some images from Yale face database are shown in Fig. 3.

As the previous mentioned, we also repeat the experiment 20 times. Table 4 shows the average recognition rate obtained by different methods with varying number of training samples and reveals that DV-KPCA is comparable to PCA and DV-PCA methods. The recognition rate of PCA, DV-PCA, and DV-KPCA with varying dimension of feature vectors for five training samples is given in Fig. 4. It can be easily ascertained from Fig. 4 that DV-KPCA, with reduced feature vector, obtains the same or even good recognition rate when compared with other methods.

Table 4
Average recognition rate (%) comparison of different approaches on Yale database.

Method/number of training samples	2	4	6	8
PCA	76.67	89.67	94.00	99.11
DV-PCA	80.22	92.33	94.56	99.22
DV-KPCA	82.19	93.12	95.13	99.44

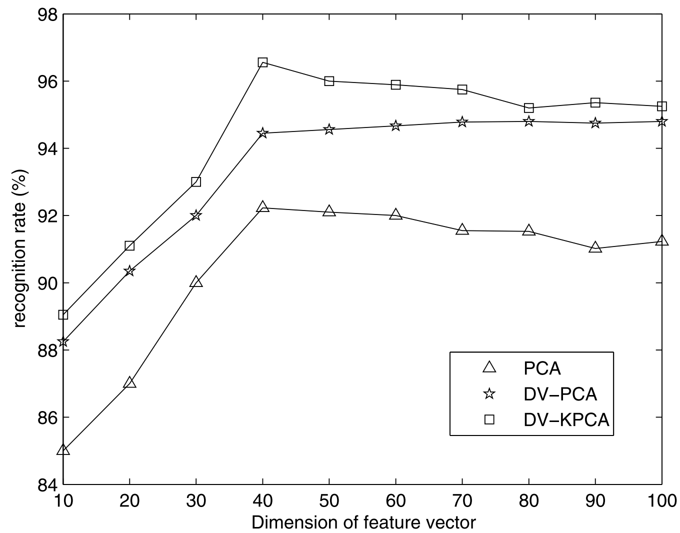

Fig. 4. Recognition performance of different approaches with varying dimension of feature vectors on Yale database.

3.3. Experiments on FERET database

We analyze the performance of the proposed DV-KPCA method using FERET database. Comparative results are also given for other methods such as KPCA, PCA and LDA. FERET database is a standard test set for face recognition technologies [27]. 450 frontal face images corresponding to 150 subjects are extracted from the database for the experiments. The face images are acquired under varying illumination conditions and facial expressions. The following procedures are applied to normalize the face images prior to the experiments:

(1) The centers of the eyes of each image are manually marked, each image is then rotated and scaled to align the centers of the eyes,
(2) Each face image is cropped to the size of 128×28 to extract facial region.

Fig. 5 shows some images from FERET database. The images are captured at different photo sessions so that they display different illumination and facial expressions. To test the algorithms, two images of each subject are randomly chosen for the training, while the remaining one is used for the testing. The experiments are repeated 3 times. The performance of the proposed method is compared with PCA, KPCA and LDA. The average results of the experiments are shown in Table 5. It can be observed that the proposed method has improved the performance of PCA and LDA significantly.

Fig. 5. Normalized face images on FERET face database.

Table 5
Recognition rate (%) comparison of different approaches on FERET database.

Method	PCA	KPCA	LDA	DV-PCA	DV-KPCA
Recognition rate (%)	42.3	48.6	69.7	70.6	74.9

Fig. 6. Original and normalized face images on AR face database.

3.4. Experiments on AR database

This face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains over 4000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. We select 1300 images of 50 males and 50 females, and each person has 13 images to test our method. The original images are of 768 by 576 pixels, and then normalized to 128×128 (shown in Fig. 6). We randomly select $5, 6, 7, 8$ samples from each class for the training, and the rest for the testing. The processes are repeated 5 times, and the results are shown in Table 6.

Table 6 reveals that DV-KPCA is comparable to PCA, LDA and DV-PCA methods in terms of recognition rate and the number of the training samples. From the table, it can be easily ascertained, with increased training number, the proposed method obtains good recognition rate when compared with other methods.

Table 6
Recognition rate (%) comparison of different approaches on AR database.

Method/number of the training sample	PCA	LDA	DV-PCA	DV-KPCA
5	65.57	75.14	84.62	88.47
6	73.39	78.23	89.58	91.45
7	78.38	85.48	90.75	92.56
8	80.21	89.57	92.49	93.78

4. Conclusion

In this paper, an efficient face recognition approach based on difference vector plus KPCA is proposed. Difference vector is gotten by the difference between the original image and common vector of its class. By KPCA procedure of the difference vectors, feature vectors are obtained. The recognition result is obtained by finding the minimum distance between the difference feature vectors of the test image and the training image. Experiments on ORL, Yale, FERET and AR database indicate that the proposed approach is encouraging.

Acknowledgments

Project was supported by The National Natural Science Foundation of China (Grant No. 60805007) and The Key Project of Ministry of Education (Grant No. 109057).

Appendix A

Theorem. *Common vector of the ith class* $x^i_{common} = x^i_k - s^i$ $(k = 1, \ldots, N)$ *(in Eq.* (3)*) is independent from a selected sample* x^i_k*, i.e.,*

$$x^i_{common-k} = x^i_{common-h}, \quad k, h = 1, \ldots, N.$$

Proof. (The definition of b_k and B in the paper.) Since $b_k \in B$ and $\{y^i_1, \ldots, y^i_{N-1}\}$ is a basis of B, there exist constants $K_1, \ldots, K_{N-1}$ such that

$$b_k = K_1 y^i_1 + \cdots + K_{N-1} y^i_{N-1} \tag{A.1}$$

If both sides of Eq. (A.1) is scalarly multiplied by each vector $\{y^i_1, \ldots, y^i_{N-1}\}$ respectively and using $\langle y^i_h, y^i_k \rangle = \delta_{kh} = \{1, \text{ if } k = h;\ 0, \text{ if } k \neq h\}$, we obtain $K_h = \langle b_k, y^i_h \rangle (h = 1, \ldots, N-1)$. Thus, the equalities

$$b_k = \langle b_k, y^i_1 \rangle y^i_1 + \cdots + \langle b_k, y^i_{N-1} \rangle y^i_{N-1} \tag{A.2}$$

or

$$x^i_k - x_1 = \langle x^i_k - x_1, y^i_1 \rangle y^i_1 + \cdots + \langle x^i_k - x_1, y^i_{N-1} \rangle y^i_{N-1} \tag{A.3}$$

can be written.

From Eq. (A.3), the following equality is written:

$$\begin{aligned} & x^i_k - \left[\langle x^i_k, y^i_1 \rangle y^i_1 + \cdots + \langle x^i_k, y^i_{N-1} \rangle y^i_{N-1}\right] \\ & \quad = x^i_1 - \left[\langle x^i_1, y^i_1 \rangle y^i_1 + \cdots + \langle x^i_1, y^i_{N-1} \rangle y^i_{N-1}\right], \\ & \quad k = 1, \ldots, N \end{aligned} \tag{A.4}$$

According to the above notation, it is clear that $x^i_{common-k} = x^i_{common-1}$, $\forall k = 1, \ldots, N$.

Hence, $x^i_{common-k} = x^i_{common-h}$, $k, h = 1, \ldots, N$. $\square$

References

[1] S.W. Lee, P.S.P. Wang, S.N. Yanushkevich, S.W. Lee, Noniterative 3D face reconstruction based on photometric stereo, International Journal of Pattern Recognition and Artificial Intelligence Biometrics 22 (3) (2008) 389–410.

[2] F.Y. Shih, S. Cheng, C.F. Chuang, P.S.P. Wang, Extracting faces and facial features from color images, International Journal of Pattern Recognition and Artificial Intelligence Biometrics 22 (3) (2008) 445–459.

[3] Y. Wen, L.H. He, An improved locally linear embedding for sparse data sets, International Journal of Pattern Recognition and Artificial Intelligence 25 (5) (2011) 763–775.

[4] Y. Wen, Y. Lu, P.F. Shi, Handwritten Bangla numeral recognition system and its application to postal automation, Pattern Recognition 40 (2007) 99–107.

[5] X.G. You, D. Zhang, et al., Face representation by using non-tensor product wavelets, in: Proc. ICPR2006, 2006, pp. 503–506.

[6] Z. He, X. You, Y.Y. Tang, P.S. Wang, Y. Xue, Texture image retrieval using novel non-separable filter banks based on centrally symmetric matrices, in: Proc. ICPR2006, 2006, pp. 161–164.

[7] C. Chatterjee, Z. Kang, V.P. Roychowdhury, Algorithms for accelerated convergence of adaptive PCA, IEEE Transactions on Neural Networks 11 (2) (2000) 338–355.

[8] B.A. Draper, K. Baek, M.S. Bartlett, J.R. Beveridgea, Recognizing faces with PCA and ICA, Computer Vision and Image Understanding 91 (2003) 115–137.

[9] Y. Wen, P.F. Shi, Image PCA: A new approach for face recognition, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2007, pp. 1241–1244.

[10] O.Y. Shan, B. Zheng, et al., Robust recursive least squares learning algorithm for principal component analysis, IEEE Transactions on Neural Networks 11 (1) (2000) 215–221.

[11] Y. Wen, P.F. Shi, A novel classifier for handwritten numeral recognition, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, pp. 1321–1324.

[12] J. Karhunen, J. Joutsensalo, Generalizations of principal component analysis, optimization problems, and neural networks, Neural Networks 8 (4) (1995) 549–562.

[13] B. Scholkopf, A. Smola, K.R. Muller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computer (1998) 1299–1319.

[14] Y. Wen, Y. Lu, et al., An algorithm for license plate recognition applied to intelligent transportation system, IEEE Transactions on Intelligent Transportation Systems 12 (3) (2011) 830–845.

[15] Z. Liang, P.F. Shi, Kernel direct discriminant analysis and its theoretical foundation, Pattern Recognition 38 (2005) 445–447.

[16] Y. Wen, P.F. Shi, An approach to numeral recognition based on improved LDA and Bhattacharyya distance, in: The Third International Symposium on Communications, Control and Signal Processing, 2008, pp. 309–311.

[17] M.B. Gulmezoglu, V. Dzhafarov, M. Keskin, A. Barkana, A novel approach to isolated word recognition, IEEE Transactions on Speech and Audio Processing 7 (6) (1999) 620–628.

[18] M.B. Gulmezoglu, V. Dzhafarov, A. Barkana, The common vector approach and its relation to principal component analysis, IEEE Transactions on Speech and Audio Processing 9 (6) (2001) 655–662.

[19] H. Cevikalp, M. Wilkes, A. Barkana, Discriminative common vectors for face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (1) (2005) 4–13.

[20] Y. He, L. Zhao, C. Zou, Kernel discriminative common vectors for face recognition, in: Proceedings of Fourth International Conference on Machine Learning and Cybernetics, 2005, pp. 18–21.

[21] Y. He, L. Zhao, C. Zou, Face recognition using common faces method, Pattern Recognition 39 (2006) 2218–2222.

[22] L.J. Cao, K.S. Chua, W.K. Chong, et al., A comparison of PCA, KPCA, and ICA for dimensionality reduction in support vector machine, Neurocomputing 55 (2003) 321–336.

[23] R. Debnath, H. Takahashi, Kernel selection for the support vector machine, IEICE Transactions on Information and System 87 (12) (2004) 2903–2904.

[24] J.S. Taylor, N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, 2004.

[25] K. Hoffman, R. Kunze, Linear Algebra, Prentice Hall, Englewood Cliffs, New Jersey, 1961.

[26] C.H. Edwards, D.E. Penney, Elementary Linear Algebra, Prentice Hall, Englewood Cliffs, New Jersey, 1988.

[27] P. Phillips, H. Moon, et al., The Feret evaluation, in: Face Recognition: From Theory to Applications, 1998, pp. 244–261.

Ying Wen received her B.Sc. degree in industry automation from Technology of Hefei University in 1997 and the M.Sc. and Ph.D. degrees in imaging processing and pattern recognition from Shanghai University and Shanghai Jiao Tong University, China, in 2002 and 2009 respectively. She is an associate professor at East China Normal University and currently does research as a Postdoctoral Research Fellow at Columbia University in the city of New York. Her research interests include documentary processing, pattern recognition, machine learning and medical image processing.

Lianghua He received his B.Sc. degree in engineering survey from Technology of Wuhan Technology of Surveying and Mapping in 1999 and his M.Sc. and Ph.D. degrees in GPS Navigation and signal and information processing from Wuhan University and Southeast University in 2002 and 2005 respectively. He currently does research as an associate professor at Tongji University. His research interests include image processing, pattern recognition, machine learning and medical image processing.

Pengfei Shi received the Bachelor's and Master's degree in electrical engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China in 1962 and 1965, respectively. In 1980, he joined the Institute of Image Processing and Pattern Recognition (IPPR), SJTU. During the past 30 years, he worked in the area of image analysis, pattern recognition and visualization. He has published more than 100 papers. He is at present the Director of the Institute of IPPR at SJTU and Professor of pattern recognition and intelligent system in the Faculty of Electronic and Information Engineering. He is a senior member of the IEEE.