IMAGE PCA: A NEW APPROACH FOR FACE RECOGNITION

Ying Wen, Pengfei Shi

Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai 200030, China

ABSTRACT

Two-dimensional principal component analysis (2DPCA) for face recognition has been proposed which is based on 2D matrices. It needs more coefficients for feature vectors than principal component analysis (PCA). In this paper, we develop an idea which is working in *the projective feature image* obtained by 2DPCA on the original images i.e., image PCA, for efficient face representation and recognition. To test image PCA and evaluate its performance, a number of experiments are performed on two face image database: ORL and Yale face databases. The experimental results show that image PCA achieves the same or even higher recognition rate than 2DPCA, while the former needs less coefficients for feature vectors than the latter.

Index Terms**—**feature extraction, image representation, face recognition

1. INTRODUCTION

Principal component analysis is a well-known feature extraction method widely used in the areas of pattern recognition, computer vision and signal processing, etc. Sirovich and Kirby [1,2] first used PCA to efficiently represent pictures of human faces. Afterwards, Turk and Pentland [3] presented the well-known Eigenfaces method for face recognition. Since then, PCA has been widely investigated and has become one of the most successful approaches in face recognition [4,5,6].

In the PCA-based face recognition methods, 2D face image matrices must be previously transformed into 1D image vectors column by column or row by row which often leads to a high-dimensional vector space, where it is difficult to evaluate the covariance matrix accurately due to its large size and the relatively small number of training samples. Furthermore, computing the eigenvectors of covariance matrix is very time-consuming. To overcome those problems, a new technique called two-dimensional principal component analysis [7] was recently proposed. As opposed to PCA, 2DPCA is based on 2D image matrices rather than 1D vector so the image matrix does not need to be transformed into a vector prior to feature extraction. Instead, an image covariance matrix is constructed directly using the original image matrix, and its eigenvectors are derived for image feature extraction. The recognition rate on several face databases was higher using 2DPCA instead of PCA, and the extraction of image features is computationally more efficient using 2DPCA but not PCA. However, the main disadvantage of 2DPCA is that it needs more *coefficients*, i.e., storage, for face recognition than PCA. Furthermore, the running time and storage depend on the number of coefficients. Daoqiang Zhang et al.[8] proposed $(2D)^2$PCA approach to solve this problem by simultaneously considering the row and column directions of the original image. In this paper, we propose another approach to solve the coefficients problem. We first obtain a family of projective feature vectors by 2DPCA on the original image, which are called *a projective feature image* of a sample image. Then we process the transpose matrix of *the projective feature image* by 2DPCA again. By considering the original image and the projective feature image, we develop the 2DPCA, i.e., image PCA (we call it IPCA here after).Experimental results on ORL and YALE face databases show that IPCA achieves the same or even higher recognition rate than 2DPCA and $(2D)^2$PCA, while the number of coefficients needed by the former for feature vectors extraction is much less than that of others. The experimental results also indicate high performance of this algorithm compared to others.

The rest of this paper is organized as follows: Section 2 briefly reviews the 2DPCA method; Section 3 presents IPCA approach; in Section 4, some experiments on ORL and Yale face databases are given to compare the performances of 2DPCA, $(2D)^2$PCA and IPCA; finally, conclusion is also given.

2. TWO-DIMENSIONAL PCA

Consider an $m\times n$ random image matrix A. Let $U \in R^{n\times d}$ be a matrix with orthonormal columns, $n\geq d$. Projecting A onto U yields an $m\times d$ matrix $Y=AU$. In 2DPCA, the total scatter of the projected samples was used to determine a good projection matrix U. That is, the following criterion is adopted

$$J(U) = trace\{E(Y-EY)(Y-EY)^T]\}$$
$$= trace\{E[(AU-E(AU))(AU-E(AU))^T]\} \quad (1)$$
$$= trace\{U^T E[(A-EA)^T(A-EA)]U]\},$$

Define the image covariance matrix $C=E[(A-EA)^T(A-EA)]$, which is an $n\times n$ nonnegative definite matrix. Suppose that there are L training face images, denoted by $m\times n$ matrices A_k $(k=1,2,...,L)$, and denote the average image as:

$$\widetilde{A} = \frac{1}{L}\sum_k A_k \quad (2)$$

Then C can be evaluated by

$$C = \frac{1}{L}\sum_{k=1}^{L}(A_k - \overline{A})^T(A_k - \overline{A}) \quad (3)$$

It has been proven that the optimal value for the projection matrix U_{opt} is composed by the orthonormal eigenvectors $U_1, \ldots, U_d$ of C corresponding to the d largest eigenvalues, i.e. $U_{opt}=[U_1, \ldots, U_d]$. Because the size of C is only $n\times n$, computing its eigenvectors is very efficient. Also, like in PCA the value of d can be controlled by setting a threshold as follows:

$$\frac{\sum_{i=1}^{d}\lambda_i}{\sum_{i=1}^{n}\lambda_i} \geq \theta \quad (4)$$

Where $\lambda_1, \lambda_2, \ldots, \lambda_n$ is the n biggest eigenvalues of C and θ is a pre-set threshold.

3. IMAGE PCA

3.1 Idea and Algorithm

For a given image sample A, the optimal projection vectors of 2DPCA, $U_1,...,U_d$, are used for feature extraction. Then, we obtain a family of projected feature vectors, $Y_1,...,Y_d$, which are called *the projective feature image* of the sample image A. *The projective feature image* are used to form an $m\times d$ matrix $B=[Y_1,...,Y_d]$. Since one disadvantage of 2DPCA (compared to PCA) is that more coefficients are needed to represent an image, we propose an approach based on the projective feature image 2DPCA to reduce the dimension of 2DPCA. The approach is called image PCA.

Consider an $m\times d$ projective feature image B (B^T is the transpose matrix B). Let $V\in R^{m\times h}$ be a matrix with orthonormal columns, $m\geq h$. Projecting B^T onto V yields a $d\times h$ matrix $Z=B^TV$. In *the projective feature image* processed by 2DPCA, the total scatter of the projected samples was used to determine a good projection matrix V. That is, the following criterion is adopted:

$$J(V) = trace\{E(Z-EZ)(Z-EZ)^T]\}$$
$$= trace\{E[(B^TV-E(B^TV))(B^TV-E(B^TV))^T]\} \quad (5)$$
$$= trace\{V^T E[(B^T-EB^T)^T(B^T-EB^T)]V]\},$$

Define *the projective feature image covariance matrix* $D=E[(B^T-E\,B^T)^T(B^T-E\,B^T)]$, which is an $m\times m$ nonnegative definite matrix. Suppose that there are L projective feature images obtained by L training face images, denoted by $m\times d$ matrices B_k $(k=1,2,...,L)$, and denote the average value as:

$$\overline{B}^T = \frac{1}{L}\sum_k B_k^T \quad (6)$$

Then D can be evaluated by

$$D = \frac{1}{L}\sum_{k=1}^{L}(B_k^T - \overline{B}^T)^T(B_k^T - \overline{B}^T) \quad (7)$$

It has been proven that the optimal value for the projection matrix V_{opt} is composed by the orthonormal eigenvectors $V_1, \ldots, V_h$ of D corresponding to the h largest eigenvalues, i.e. $V_{opt}=[V_1, \ldots, V_h]$. Also, like in PCA the value of h can be controlled by setting a threshold as follows:

$$\frac{\sum_{i=1}^{h}\alpha_i}{\sum_{i=1}^{m}\alpha_i} \geq \theta \quad (8)$$

where $\mu_1, \mu_2, \ldots, \mu_m$ is the m biggest eigenvalues of D and α is a pre-set threshold.

3.2 IPCA-Based Image Reconstruction

In the Eigenfaces method, the principal components and eigenvectors (eigenfaces) can be combined to reconstruct the image of a face. Similarly, IPCA can be used to reconstruct a face image in the following way.

Suppose the orthonormal eigenvectors corresponding to the first d largest eigenvectors of the image covariance matrix C are $U_1,...,U_d$. After the image samples are projected onto these axes, the resulting principal component vectors are $Y_k=AU_k$ $(k=1,2,...,d)$. Let $B=[Y_1,...,Y_d]$ and $U=[U_1,...,U_d]$, then

$$B=AU \quad (9)$$

Since $U_1,...,U_d$ are orthonormal, from Eq.(9), it is easy to obtain the reconstructed image of sample A:

$$\widetilde{A} = BU^T = \sum_{k=1}^{d} Y_k U_k^T \quad (10)$$

Let $\widetilde{A} = YU^T = AUU^T$, which is of the same size as image A, and represents the reconstructed subimage of A. That is, image A can be approximately reconstructed by adding up the first d subimages.

Similarly, suppose the orthonormal eigenvectors corresponding to the first h largest eigenvectors of the projective feature image covariance matrix D are $V_1, \ldots, V_h$. After *the projected feature images* are projected onto these axes, the resulting principal component vectors are $Z_k=B^TV_k$ $(k=1,2,...,h)$. Let $Z=[Z_1,...,Z_h]$ and $V=[V_1,...,V_h]$, then

$$Z=B^TV \quad (11)$$

Since $V_1,...,V_h$ are orthonormal, from Eq.(11), it is easy to obtain the reconstructed image of sample B^T:

$$\widetilde{B}^T = ZV^T = \sum_{k=1}^{h} Z_k V_k^T \quad (12)$$

Let $\widetilde{B}^T = Z\,V^T = B^TV\,V^T = (AU)^T VV^T$, which is of the same size as projected feature image B^T, and represents the reconstructed projective feature subimage of B^T. That is, projective feature image B^T can be approximately reconstructed by adding up the first h projective feature subimages.

Based on the above statement, the reconstructed image can be obtained by

$$\widetilde{A} = AUU^T = (B^T)^T U^T = (\widetilde{B}^T)^T U^T \\ = ((AU)^T VV^T)^T U^T = VV^T AUU^T \tag{13}$$

In particular, when the selected number of principal component vectors $d=n$ (n is the total number of eigenvectors of C), $h=m$ (m is the total number of eigenvectors of D), we have $\widetilde{A} = A$, i.e., the image is completely reconstructed without any loss of information. Otherwise, if $d<n$ and $h<m$, the reconstructed image $\widetilde{A}$ is an approximation for A.

3.3 Feature Extraction and Classification

The optimal projection vectors of IPCA, $U_1,\ldots,U_d$, $V_1, \ldots ,V_h$, are used for feature extraction. For a given image sample A, let

$$Q = B^T V = U^T A^T V \tag{14}$$

Then, we obtain a family of projected feature vectors, i.e. $d\times h$ matrix Q. When used for face recognition, the matrix Q is also called the feature matrix. After projecting each training image A_k $(k=1,2,\ldots,L)$ onto U and V, we obtain the training feature matrices Q_k $(k=1,2,\ldots,L)$. Given a test face image A, first use Eq.(14) to get the feature matrix Q, then a nearest neighbor classifier is used for classification. Here the distance between Q and Q_k is defined by

$$d(Q,Q_k) = \|Q - Q_k\| = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{h}(Q^{(i,j)} - Q_k^{(i,j)})^2} \tag{15}$$

4. EXPERIMENTS AND ANALYSIS

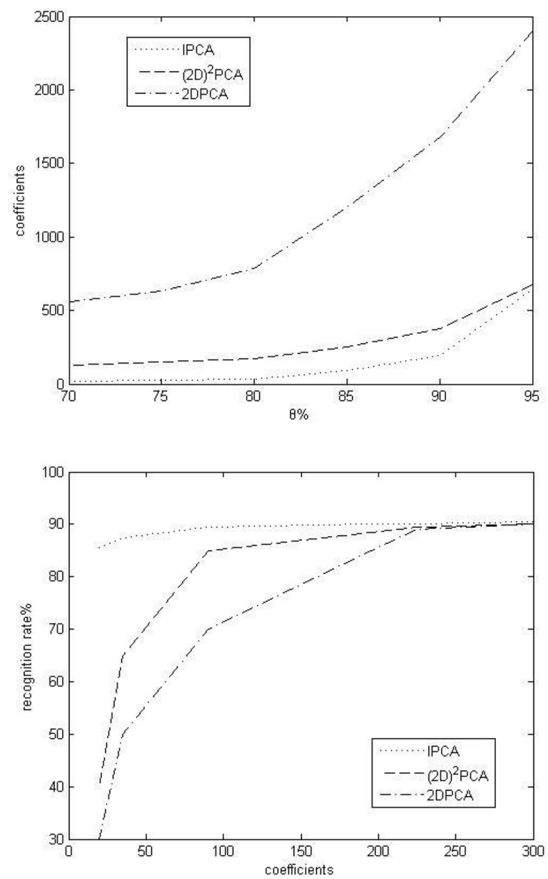
The proposed IPCA approach is used for face recognition and tested on two well-known face image databases (ORL and Yale). If without extra explanations, α (in Eq.(8)) is set to 0.999 to reserve the most information of the projective feature image in this paper. All of our experiments are carried out on a PC machine with P4 1.6GHz CPU and 256MB memory.

4.1 Experiments on ORL Database

The ORL database (http://www.cam-orl.co.uk) contains images from 40 individuals, each providing 10 different images. All images are grayscale and normalized to a resolution of 112×92 pixels. First, an experiment is performed using the first five image samples per class for training, and the remaining images for test. Thus, the total number of training samples and testing samples are both 200. Table 1 gives the comparisons of four approaches on recognition rate, dimensions of feature vector and running times. In this experiment, the number of projection vectors

Table 1 Comparisons of four approaches on ORL database

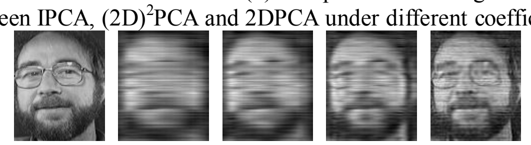
Approach	Recognition Rate(%)	Dimension	Time(s)
PCA	88	110	36.5
2DPCA	90	112×27	8.78
$(2D)^2$PCA	90	25×27	4.34
IPCA	90	27×24	4.21



(a)

(b)

Fig.1. (a) Comparisons of coefficients between IPCA, $(2D)^2$PCA and 2DPCA under different θ. (b) Comparisons of recognition rate between IPCA, $(2D)^2$PCA and 2DPCA under different coefficients.



A face image $\theta=0.7$ $\theta=0.8$ $\theta=0.9$ $\theta=0.95$

Fig.2. Some reconstructed images by IPCA.

in all approaches is controlled by the value of θ (in Eq.(4)), which is set to 0.95. Table 1 shows that 2DPCA, $(2D)^2$PCA and IPCA achieve the same improvements in recognition rate than PCA on this database, while the latter needs much reduced dimension of feature vector for the following classification than the former two. Table 1 also indicates that IPCA needs the least running time among the four apporaches. It can be seen that the running time depends on dimension, i.e., storage. In this paper, dimension we called is defined as *coefficients*.

Fig.3. Some reconstructed images on ORL database. First row: original images. Second row: images gotten by 2DPCA. Third row: images gotten by $(2D)^2$ PCA. Bottom row: images gotten by the proposed method.

To further disclose the relationship between the recognition rate, *coefficients* and θ, classification experiments under different coefficients between IPCA, 2DPCA and $(2D)^2$PCA are performed and the results are plotted in Fig.1 (a) and (b), respectively. It can be seen from Fig.1 (a) that under the same θ, IPCA obtains less *coefficients* than both 2DPCA and $(2D)^2$PCA and from Fig.1 (b) that under the same *coefficients*, IPCA obtains better recognition rate than both 2DPCA and $(2D)^2$PCA. Fig.2 shows some reconstructed images of a face image in the ORL face database. The reconstructed image becomes clear as θ increases. For comparison, the 2DPCA and $(2D)^2$ PCA are also used to represent and reconstruct the same face image. Fig.3 also shows the reconstructed images under similar coefficients. It can be shown that IPCA yields higher quality images than the other two methods, when using similar amount of storage.

4.2 Experiments on Yale Database

The last experiment is performed using the Yale face database, which contains *165* images of *15* individuals (each person has 11 different images) under various facial expressions and lighting conditions. Each image is manually cropped and resized to 100×80 pixels in this experiment showed. In this experiment, we use the first k samples per class for training and the remaining samples for testing. Table 2 presents the recognition rate and *coefficients* using 2DPCA, $(2D)^2$PCA, and IPCA which correspond to different numbers of training samples. It can be shown that IPCA possesses less *coefficients* than the other two methods, when acquiring the similar recognition rate.

5. CONCLUSION

In this paper, an efficient face recognition approach called IPCA is proposed. Based on the theory of 2DPCA, we process the original image to obtain the projective feature image which then is processed by 2DPCA. We complete face recognition by feature extraction according to the coefficients obtained by IPCA on the face image. The main advantage of IPCA over 2DPCA is that the coefficients needed by the former are less than the latter while achieving higher or similar accuracy. The experimental results show that the proposed approach is effective for face recognition.

ACKNOWLEDGEMENTS

This research is supported by the national 863 project No.2006AA01Z119.

REFERENCE

[1] L. Sirovich and M. Kirby, "Low-Dimensional Procedure for Characterization of Human Faces," J. Optical Soc. Am., vol.4, pp.519-524, 1987.

[2] M. Kirby and L. Sirovich, "Application of the KL Procedure for the Characterization of Human Faces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.12, no.1, pp.103-108, Jan.1990.

[3] M. Turk and A. Pentland, "Eigenfaces for Recognition," J. Cognitive Neuroscience, vol.3, no.1, pp. 71-86, 1991.

[4] Issam Dagher and Rabih Nachar, "Face Recognition Using IPCA-ICA Algorithm", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.28, no.6, pp.996-1000, Jun 2006

[5] Chengjun Liu, "Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.26, no.5, pp.572-581,May2004.

[6] Chengjun Liu, "Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.26, no.5, pp.572-581,May2004.

[7] J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, "Two-dimensional PCA: a new approach to appearance based face representation and recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.26, no.1,pp.131–137, Jan 2004.

[8] Daoqiang Zhang, Zhi-Hua Zhou, "$(2D)^2$PCA: Two-directional two-dimensional PCA for efficient face representation and recognition", Neurocomputing 69, pp.224-231, 2005.

Table 2 Comparison of the recognition rate(%) and coefficients of three approaches on Yale face database

Approach	Training Samples (the first k samples per class)				
	3	4	5	6	7
2DPCA	69.2 (100×13)	75.2 (100×9)	78.2 (100×13)	82.7 (100×12)	83.4 (100×12)
$(2D)^2$PCA	70 (16×13)	72.4 (17×9)	78.2 (18×13)	81.4 (17×12)	85 (16×12)
IPCA	69.2 (13×12)	75.2 (9×8)	75.5 (13×12)	84 (12×10)	85 (12×11)

The value in parentheses denotes the coefficients of feature vectors for the recognition rate