$(2D)^2$PCA: 2-Directional 2-Dimensional PCA for Efficient Face Representation and Recognition

Daoqiang Zhang[1,2] and Zhi-Hua Zhou[1*]

[1] National Laboratory for Novel Software Technology

Nanjing University, Nanjing 210093, China

[2] Department of Computer Science and Engineering

Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract

Recently, a new technique called 2-dimensional principal component analysis (2DPCA) was proposed for face representation and recognition. The main idea behind 2DPCA is that it is based on 2D matrices as opposed to the standard PCA, which is based on 1D vectors. Although 2DPCA obtains higher recognition accuracy than PCA, a vital unresolved problem of 2DPCA is that it needs many more coefficients for image representation than PCA. In this paper, we first indicate that 2DPCA is essentially working in the row direction of images, and then propose an alternative 2DPCA which is working in the column direction of images. By simultaneously considering the row and column directions, we develop the 2-Directional 2DPCA, i.e. $(2D)^2$PCA, for efficient face representation and recognition. Experimental results on ORL and a subset of FERET face databases show that $(2D)^2$PCA achieves the same or even higher recognition accuracy than 2DPCA, while the former needs a much reduced coefficient set for image representation than the latter.

Keywords: Principal component analysis (PCA); Eigenface; 2-Dimensional PCA; image representation; face recognition

* Corresponding author. Email: zhouzh@nju.edu.cn, Tel.: +86-25-8368-6268, fax: +86-25-8368-6268.

1 Introduction

Principal component analysis (PCA) [3] is a well-known feature extraction and data representation technique widely used in the areas of pattern recognition, computer vision and signal processing, etc [2], [4]. In the PCA-based face representation and recognition methods [6], [9], [10], the 2D face image matrices must be previously transformed into 1D image vectors column by column or row by row. However, concatenating 2D matrices into 1D vectors 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 [7]. Furthermore, computing the eigenvectors of a large size covariance matrix is very time-consuming.

To overcome those problems, a new technique called 2-dimensional principal component analysis (2DPCA) [7] was recently proposed, which directly computes eigenvectors of the so-called *image covariance matrix* without matrix-to-vector conversion. Because the size of the image covariance matrix is equal to the width of images, which is quite small compared with the size of a covariance matrix in PCA, 2DPCA evaluates the image covariance matrix more accurately and computes the corresponding eigenvectors more efficiently than PCA. It was reported in [7] that the recognition accuracy on several face databases was higher using 2DPCA than PCA, and the extraction of image features is computationally more efficient using 2DPCA than PCA.

However, the main disadvantage of 2DPCA is that it needs many more coefficients for image representation than PCA [7], [8]. For example, suppose the image size is 100×100, then the number of coefficients of 2DPCA is 100×d, where d is usually set to no less than 5 for satisfying accuracy. Although this problem can be alleviated by using PCA after 2DPCA for further dimensional reduction, it is still unclear how the dimension of 2DPCA could be reduced directly [7]. In this paper, we first indicate that 2DPCA is essentially working in the row direction of images, and then propose an alternative 2DPCA which is working in the column direction of images. By simultaneously considering the row and column directions (see Eq. (8) in Section 4) , we develop the 2-Directional 2DPCA, i.e. $(2D)^2$PCA, for efficient face representation and recognition. Experimental results on ORL and a subset of FERET face databases show that $(2D)^2$PCA achieves the same or even higher recognition accuracy than 2DPCA, while the number

of coefficients needed by the former for image representation is much less than that of the latter. The experimental results also indicate that $(2D)^2$PCA is more computationally efficient than both PCA and 2DPCA.

The rest of this paper is organized as follows: Section 2 briefly reviews the 2DPCA method; Section 3 presents an alternative 2DPCA method; The proposed $(2D)^2$PCA method is introduced in Section 4; In Section 5, some experiments on several face databases are given to compare the performances of 2DPCA and $(2D)^2$PCA; Finally, we conclude in Section 6.

2 2DPCA

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

$$\begin{aligned} J(\mathbf{X}) &= trace\left\{ E\left[(\mathbf{Y} - E\mathbf{Y})(\mathbf{Y} - E\mathbf{Y})^T \right] \right\} \\ &= trace\left\{ E\left[(\mathbf{AX} - E(\mathbf{AX}))(\mathbf{AX} - E(\mathbf{AX}))^T \right] \right\}, \\ &= trace\left\{ \mathbf{X}^T E\left[(\mathbf{A} - E\mathbf{A})^T (\mathbf{A} - E\mathbf{A}) \right] \mathbf{X} \right\} \end{aligned} \tag{1}$$

where the last term in Eq. (1) results from the fact that trace(***AB***)=trace(***BA***), for any two matrices [1]. Define the *image covariance matrix* $\mathbf{G} = E\left[(\mathbf{A} - E\mathbf{A})^T (\mathbf{A} - E\mathbf{A}) \right]$, which is an *n* by *n* nonnegative definite matrix. Suppose that there are *M* training face images, denoted by *m* by *n* matrices $\mathbf{A}_k (k = 1, 2, ...M)$, and denote the average image as $\bar{\mathbf{A}} = \frac{1}{M} \sum_k \mathbf{A}_k$. Then $\mathbf{G}$ can be evaluated by

$$\mathbf{G} = \frac{1}{M} \sum_{K=1}^{M} (\mathbf{A}_k - \bar{\mathbf{A}})^T (\mathbf{A}_k - \bar{\mathbf{A}}). \tag{2}$$

It has been proven that the optimal value for the projection matrix $\mathbf{X}_{opt}$ is composed by the orthonormal eigenvectors $\mathbf{X}_1, ..., \mathbf{X}_d$ of $\mathbf{G}$ corresponding to the *d* largest eigenvalues, i.e. $\mathbf{X}_{opt} = [\mathbf{X}_1, ..., \mathbf{X}_d]$. Because the size of $\mathbf{G}$ is only *n* by *n*, computing its eigenvectors is very efficient. Also, like in PCA the value of *d* can be controlled by setting a threshold as follows

$$\frac{\sum_{i=1}^{d}\lambda_i}{\sum_{i=1}^{n}\lambda_i} \geq \theta \tag{3}$$

where $\lambda_1, \lambda_2, ..., \lambda_n$ is the *n* biggest eigenvalues of $\mathbf{G}$ and θ is a pre-set threshold.

3 Alternative 2DPCA

Let $\mathbf{A}_k = [(\mathbf{A}_k^{(1)})^T (\mathbf{A}_k^{(2)})^T ... (\mathbf{A}_k^{(m)})^T]^T$ and $\bar{\mathbf{A}} = [(\bar{\mathbf{A}}^{(1)})^T (\bar{\mathbf{A}}^{(2)})^T ... (\bar{\mathbf{A}}^{(m)})^T]^T$, where $\mathbf{A}_k^{(i)}$ and $\bar{\mathbf{A}}^{(i)}$ denote the *i*-th row vectors of $\mathbf{A}_k$ and $\bar{\mathbf{A}}$ respectively. Then Eq. (2) can be rewritten as

$$\mathbf{G} = \frac{1}{M}\sum_{K=1}^{M}\sum_{i=1}^{m}(\mathbf{A}_k^{(i)} - \bar{\mathbf{A}}^{(i)})^T (\mathbf{A}_k^{(i)} - \bar{\mathbf{A}}^{(i)}). \tag{4}$$

Equation (4) reveals that the image covariance matrix $\mathbf{G}$ can be obtained from the outer product of row vectors of images, assuming the training images have zero mean, i.e. $\bar{\mathbf{A}} = (0)_{m\times n}$. For that reason, we claim that original 2DPCA is working in the row direction of images.

Illuminated by Eq. (4), a natural extension is to use the outer product between column vectors of images to construct $\mathbf{G}$. Let $\mathbf{A}_k = [(\mathbf{A}_k^{(1)})(\mathbf{A}_k^{(2)})...(\mathbf{A}_k^{(m)})]$ and $\bar{\mathbf{A}} = [(\bar{\mathbf{A}}^{(1)})(\bar{\mathbf{A}}^{(2)})...(\bar{\mathbf{A}}^{(m)})]$, where $\mathbf{A}_k^{(j)}$ and $\bar{\mathbf{A}}^{(j)}$ denote the *j*-th column vectors of $\mathbf{A}_k$ and $\bar{\mathbf{A}}$ respectively. Then an alternative definition for image covariance matrix $\mathbf{G}$ is:

$$\mathbf{G} = \frac{1}{M}\sum_{k=1}^{M}\sum_{j=1}^{n}(\mathbf{A}_k^{(j)} - \bar{\mathbf{A}}^{(j)})(\mathbf{A}_k^{(j)} - \bar{\mathbf{A}}^{(j)})^T. \tag{5}$$

Now we will show how Eq. (5) can be derived at a similar way as in 2DPCA. Let $\mathbf{Z} \in R^{m\times q}$ be a matrix with orthonormal columns. Projecting the random matrix ***A*** onto ***Z*** yields a *q* by *n* matrix $\mathbf{B} = \mathbf{Z}^T\mathbf{A}$. Similar as in Eq. (1), the following criterion is adopted to find the optimal projection matrix ***Z***:

$$\begin{aligned} J(\mathbf{Z}) &= trace\left\{E\left[(\mathbf{B} - E\mathbf{B})(\mathbf{B} - E\mathbf{B})^T\right]\right\} \\ &= trace\left\{E\left[(\mathbf{Z}^T\mathbf{A} - E(\mathbf{Z}^T\mathbf{A}))(\mathbf{Z}^T\mathbf{A} - E(\mathbf{Z}^T\mathbf{A}))^T\right]\right\}, \\ &= trace\left\{\mathbf{Z}^T E\left[(\mathbf{A} - E\mathbf{A})(\mathbf{A} - E\mathbf{A})^T\right]\mathbf{Z}\right\} \end{aligned} \tag{6}$$

From Eq. (6), the alternative definition of image covariance matrix $\mathbf{G}$ is:

$$\begin{aligned}\mathbf{G} &= E\left[(\mathbf{A}-E\mathbf{A})(\mathbf{A}-E\mathbf{A})^T\right] = \frac{1}{M}\sum_{k=1}^{M}(\mathbf{A}_k-\bar{\mathbf{A}})(\mathbf{A}_k-\bar{\mathbf{A}})^T \\ &= \frac{1}{M}\sum_{k=1}^{M}\sum_{j=1}^{n}(\mathbf{A}_k^{(j)}-\bar{\mathbf{A}}^{(j)})(\mathbf{A}_k^{(j)}-\bar{\mathbf{A}}^{(j)})^T\end{aligned} . \quad (7)$$

Similarly, the optimal projection matrix $\mathbf{Z}_{opt}$ can be obtained by computing the eigenvectors $\mathbf{Z}_1,...,\mathbf{Z}_q$ of Eq. (7) corresponding to the q largest eigenvalues, i.e. $\mathbf{Z}_{opt}=[\mathbf{Z}_1,...,\mathbf{Z}_q]$. The value of q can also be controlled by setting a threshold as in Eq. (3). Because the eigenvectors of Eq. (7) only reflect the information between columns of images, we say that the alternative 2DPCA is working in the column direction of images.

4 (2D)^{2}PCA

As discussed in Section 2 and Section 3, 2DPCA and alternative 2DPCA only works in the row and column direction of images respectively. That is, 2DPCA learns an optimal matrix ***X*** from a set of training images reflecting information between rows of images, and then projects an m by n image ***A*** onto ***X***, yielding an m by d matrix $\mathbf{Y}=\mathbf{AX}$. Similarly, the alternative 2DPCA learns optimal matrix ***Z*** reflecting information between columns of images, and then projects ***A*** onto ***Z***, yielding a q by n matrix $\mathbf{B}=\mathbf{Z}^T\mathbf{A}$. In the following, we will present a way to simultaneously use the projection matrices ***X*** and ***Z***.

Suppose we have obtained the projection matrices ***X*** (in Section 2) and ***Z*** (in Section 3), projecting the m by n image ***A*** onto ***X*** and ***Z*** simultaneously, yielding a q by d matrix ***C***

$$\mathbf{C}=\mathbf{Z}^T\mathbf{AX}. \quad (8)$$

The matrix ***C*** is also called the coefficient matrix in image representation, which can be used to reconstruct the original image ***A***, by

$$\hat{\mathbf{A}}=\mathbf{ZCX}^T. \quad (9)$$

When used for face recognition, the matrix ***C*** is also called the feature matrix. After projecting each training image $\mathbf{A}_k(k=1,2,...M)$ onto ***X*** and ***Z***, we obtain the training feature matrices $\mathbf{C}_k(k=1,2,...M)$. Given a test face image ***A***, first use Eq. (8) to get the feature matrix ***C***, then a

nearest neighbor classifier is used for classification. Here the distance between $\boldsymbol{C}$ and $\mathbf{C}_k$ is defined by

$$d(\mathbf{C},\mathbf{C}_k)=\left\|\mathbf{C}-\mathbf{C}_k\right\|=\sqrt{\sum_{i=1}^{q}\sum_{j=1}^{d}\left(C^{(i,j)}-C_k^{(i,j)}\right)^2} \tag{10}$$

5 Experiments

In this section, we experimentally evaluate our proposed $(2D)^2$PCA method with PCA, 2DPCA and alternative 2DPCA, on two well-known face databases: ORL and FERET [5]. All of our experiments are carried out on a PC machine with P4 1.7GHz CPU and 256MB memory. If without extra explanations, the number of projection vectors in all methods are controlled by the value of θ, which is set to 0.95 in all the experiments.

5.1 Results on ORL database

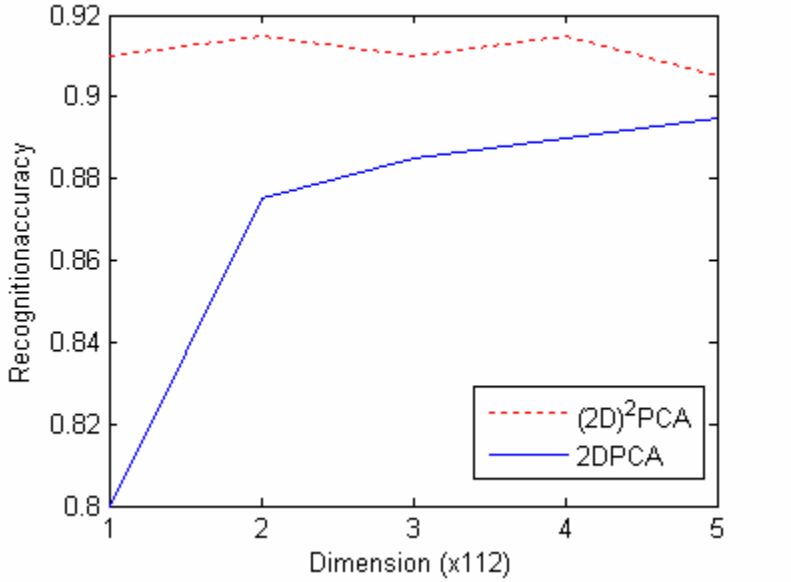
The ORL database (http://www.uk.research.att.com/facedatabase.html) contains images from 40 individuals, each providing 10 different images with size of 112×92. In this experiment, the first five image samples per class are used for training, and the remaining images for test. Table 1 gives the comparisons of four methods on recognition accuracy, dimensions of feature vector and running times. Table 1 shows that 2DPCA, alternative 2DPCA and $(2D)^2$PCA achieves the same improvements in accuracy 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 $(2D)^2$PCA needs the least running time among the four methods.

Table 1. Comparisons of four methods on ORL database.

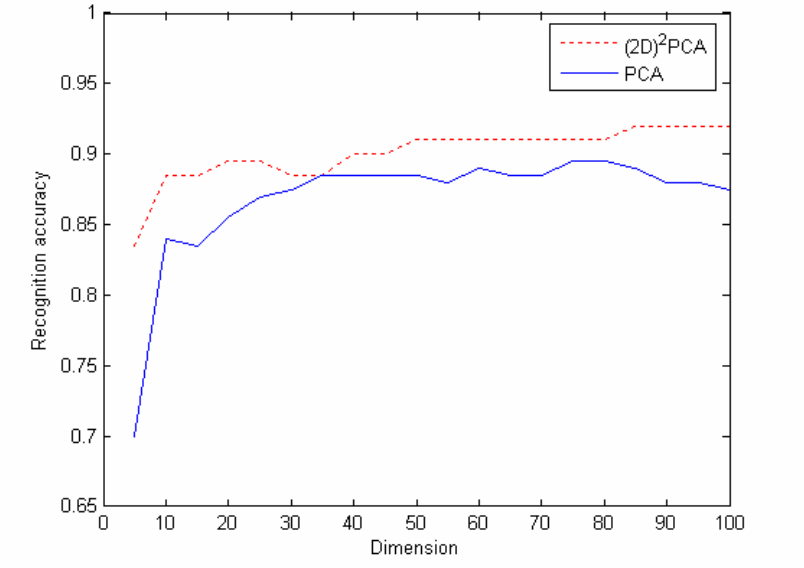
Method	Accuracy (%)	Dimension	Time (s)
PCA	88.0	110	26.65
2DPCA	90.5	27×112	7.30
Alternative 2DPCA	90.5	26×92	5.77
$(2D)^2$PCA	90.5	27×26	3.43

To further disclose the relationship between the accuracy and dimension of feature vectors, classification experiments under a series of different dimensions between $(2D)^2$PCA and PCA,

2DPCA are performed and the results are plotted in Fig. 1(a) and 1(b) respectively. It can be seen from Figs. 1 that under the same dimensions of feature vectors, $(2D)^2$PCA obtains better accuracy than both 2DPCA and PCA.



(a)



(b)

Figure 1 Comparisons of accuracies between $(2D)^2$PCA and 2DPCA (a), and between $(2D)^2$PCA and PCA (b) under different dimensions.

5.2 Results on partial FERET database

This partial FERET face database comprises 400 gray-level frontal view face images from 200 persons, each of which is cropped with the size of 60×60. There are 71 females and 129 males; each person has two images (**fa** and **fb**) with different facial expressions. The **fa** images are used as gallery for training while the **fb** images as probes for test. Table 2 gives the comparisons of four methods on recognition accuracy, dimensions of feature vector and running times. Also, $(2D)^2$PCA outperforms the other methods in accuracy and speed.

Table 2. Comparisons of four methods on partial FERET database.

Method	Accuracy (%)	Dimension	Time (s)
PCA	83.0	73	10.32
2DPCA	84.5	13×60	1.82
Alternative 2DPCA	84.5	14×60	1.80
$(2D)^2$PCA	85.0	13×14	1.15

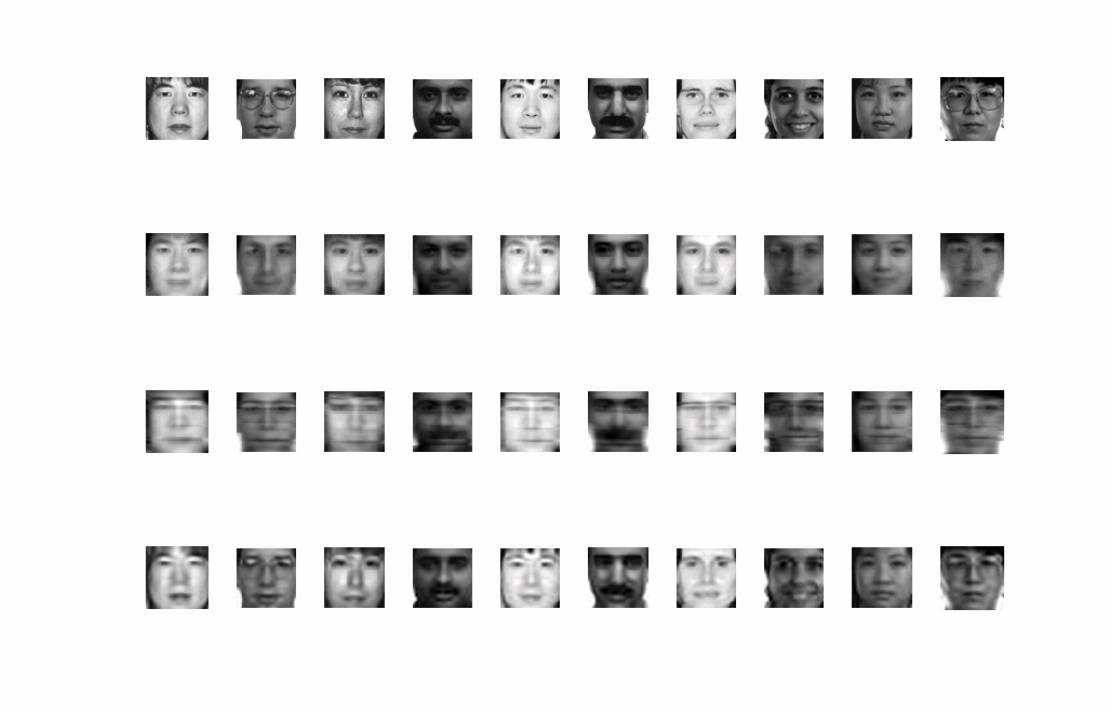


Figure 2 Some reconstructed training images on FERET database. First row: original images. Second row: images gotten by PCA. Third row: images gotten by 2DPCA. Bottom row: images gotten by the proposed **$(2D)^2$PCA** method.

Finally, experiments are carried out to compare abilities of PCA, 2DPCA and $(2D)^2$PCA in representing face images under similar compression ratios. Suppose there are M m by n training face images, the number of projection vectors in PCA, 2DPCA and alternative 2DPCA is p, d and q. Then the compression ratios of PCA, 2DPCA, alternative 2DPCA and $(2D)^2$PCA are computed as $Mmn/(Mp+mnp)$, $Mmn/(Mmd+nd)$, $Mmn/(Mnq+mq)$ and $Mmn/(Mdq+nd+mq)$ respectively. Figure 2 plots some reconstructed training face images using PCA, 2DPCA and $(2D)^2$PCA on FERET database under similar compression ratios. It can be shown that $(2D)^2$PCA yields higher quality images than the other two methods, when using similar amount of storage.

6 Conclusions

In this paper, an efficient face representation and recognition method called $(2D)^2$PCA is proposed. The main difference between $(2D)^2$PCA and existing 2DPCA is that the latter only works in the row direction of face images, while the former works simultaneously in the row and the column directions of face images. The main advantage of $(2D)^2$PCA over 2DPCA lies in that the number of coefficients needed by the former for face representation and recognition is much smaller than the latter. Experimental results show the effects of the proposed method.

Note that Eqs. (8) and (9) are essentially different from singular value decomposition (SVD). On one hand, the $\boldsymbol{C}$ here is not a diagonal matrix as that used in SVD. On the other hand, in contrast to SVD where the $\boldsymbol{X}$ and $\boldsymbol{Z}$ are only relative to the current matrix $\boldsymbol{A}$, here the $\boldsymbol{X}$ and $\boldsymbol{Z}$ are computed beforehand and the columns of them correspond to the eigenvectors of Eqs. (2) and (5) respectively.

Acknowledgements

The authors would like to thank the anonymous referees for their helpful comments and suggestions. This work was supported by the Postdoctoral Science Foundation of China, Jiangsu Postdoctoral Research Foundation, the National Science Foundation of China under Grant No. 60496320, and the Jiangsu Science Foundation under Grant No. BK2004001. Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

References

[1] R. Bellman, *Introduction to Matrix Analysis*, 2nd Edition, McGraw-Hill, NY, USA, pp. 96, 1970.

[2] A.K. Jain, R.P.W Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4-37, Jan. 2000.

[3] I.T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.

[4] 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.

[5] P.J. Phillips, H. Wechsler, J. Huang, and P.J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol.16, no.5, pp. 295-306, Apr. 1998.

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

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

[8] D.Q. Zhang, S.C. Chen, and J. Liu, "Representing image matrices: Eigenimages vs. Eigenvectors," In: *Proceedings of the 2nd International Symposium on Neural Networks (ISNN'05)*, Chongqing, China, vol. 2, pp. 659-664, 2005.

[9] L. Zhao and Y. Yang, "Theoretical analysis of illumination in PCA-based vision systems," *Pattern Recognition*, vol. 32, no. 4, pp.547-564, Apr. 1999.

[10] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips. Face recognition: a literature survey. http://citeseer.nj.nec.com/374297.html, 2000.