Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

$(2D)^2$PCA-ICA: A New Approach for Face Representation and Recognition

Dongmin Jeong, Minho Lee
Department of Electrical Engineering and Computer Science
Kyungpook National University
1370 Sankyuk-Dong, Puk-Gu, Taegu 702-701, Korea
dmjeong@ee.knu.ac.kr, mholee@knu.ac.kr

Sang-Woo Ban
Department of Information and Communication Engineering
Dongguk University
707 Seokjang-Dong, Gyeongju, Gyeongbuk 780-714, Korea
swban@dongguk.ac.kr

***Abstract*— In this paper, a new feature extraction algorithm considering both two-directional two-dimensional principal component analysis ($(2D)^2$PCA) and independent component analysis(ICA), called $(2D)^2$PCA-ICA, is proposed for face representation. This algorithm analyzes the principal components of image vectors on 2D matrices by simultaneously considering the row and column directions as opposed to the standard PCA based on 1D vectors, and transforming those principal components to the independent components that maximize the non-Gaussianity of the sources. These two major techniques such as $(2D)^2$PCA and ICA are used sequentially in order to obtain the most efficient features that properly describe a whole set of human faces in face databases. The proposed algorithm is applied to the face recognition problem. Simulation results on ORL and Yale B face databases shows that the proposed algorithm achieves high average success rate in face recognition compared with other models.**

***Keywords*—2DPCA, $(2D)^2$PCA, ICA, face recognition**

I. INTRODUCTION

Various face recognition techniques use face representations obtained by unsupervised statistical methods. Typically, the face representation methods find a set of basis images and represent faces as a linear combination of those basis images. For the same purpose, this paper sequentially merges two techniques of two-directional two-dimensional principal component analysis ($(2D)^2$PCA) and independent component analysis (ICA).

The PCA is a well-known feature extraction and data representation technique widely used in the areas of pattern recognition, computer vision and signal processing, etc [1]. Sirovich and Kirby [2, 3] firstly used the PCA to efficiently represent human face pictures. Afterwards, Turk and Pentland [4] introduced the well-known eigenfaces method for face recognition. Since then, the PCA has been widely investigated and has become one of the most successful approaches in face recognition [5, 6]. In the PCA-based face representation and recognition methods, the 2D face image matrices must be previously transformed into 1D image vectors constructed by 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 two-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. The 2DPCA evaluates a covariance matrix of an image more accurately, and computes the corresponding eigenvectors more efficiently than the conventional PCA 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 the PCA. It was reported in [7] that the higher recognition accuracy on several face databases was generated by the model using the 2DPCA than the PCA. Moreover, the extraction of image features is conducted more efficiently in computation time of view by the 2DPCA than the PCA. However, the main disadvantage of the 2DPCA is that it needs more coefficients for image representation than the PCA [7, 8]. Daoqiang Zhang et al. [9] proposed a two-directional 2DPCA, i.e. $(2D)^2$PCA approach to solve this problem by simultaneously considering the row and column directions of the original image.

In addition, the ICA algorithm is considered to estimate the independent characterization of human faces. Barlow has argued that such redundancy provides knowledge [10] and that the role of the sensory system is to develop factorial representations in which these dependencies are separated into independent components (ICs). Barlow also argued that such representations are advantageous for encoding complex objects that are characterized by high-order dependencies. Atick and Redlich have also argued for such representations as a general coding strategy for the visual system [11]. In a task such as face recognition, in which important information may be contained in the high-order relationships among pixels, it seems reasonable to expect that better basis images may be found by methods sensitive to these high-order statistics. The ICA is such a model having a property of a generalization of the PCA. It should be noted that the ICA developed by Bell and Sejnowski [12, 13] is much related to the method called blind source separation (BSS) [14], in which correlated sources are separated into uncorrelated sources without previous knowledge about the correlation between the elements of the sources.

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

Motivated by the fundamental characteristics of $(2D)^2$PCA and ICA, this paper proposes a new method considering both $(2D)^2$PCA and ICA, in which face image matrix is properly de-correlated and the dimension of facial features is reduced a lot. Experimental results on ORL and Yale B face databases show that $(2D)^2$PCA-ICA achieves higher recognition accuracy than other methods, while the number of coefficients for image representation is much less than that of the conventional methods.

The rest of this paper is organized as follows: Section 2 briefly reviews the $(2D)^2$PCA method; Section 3 briefly reviews the ICA method; the proposed $(2D)^2$PCA-ICA method is introduced in Section 4; in Section 5, some experiments on ORL and Yale B face databases are given to compare the performances of PCA, ICA, 2DPCA, PCA-ICA and $(2D)^2$PCA-ICA. Finally, we conclude and discuss about a further work in Section 6.

II. $(2D)^2$PCA

A. Two-demensional PCA (2DPCA)

Consider a m x n random image matrix X. Let $\mathrm{B} \in R^{n\times d}$ be a matrix with orthonormal columns, $n \geq d$. Projecting X onto B yields an m x d matrix $\mathrm{Y} = \mathrm{XB}$. In 2DPCA, the total scatter of the projected samples was used to determine a good projection matrix B. That is, the following criterion is adopted [7]:

$$\begin{aligned} J(\mathrm{B}) &= trace\{E[(\mathrm{Y}-E[\mathrm{Y}])(\mathrm{Y}-E[\mathrm{Y}])^{\mathrm{T}}]\} \\ &= trace\{E[(\mathrm{XB}-E[\mathrm{XB}])(\mathrm{XB}-E[\mathrm{XB}])^{\mathrm{T}}]\} \qquad (1) \\ &= trace\{\mathrm{B}^{\mathrm{T}}E[(\mathrm{X}-E[\mathrm{X}])^{\mathrm{T}}(\mathrm{X}-E[\mathrm{X}])]\mathrm{B}\}, \end{aligned}$$

where $\mathrm{G} = E[(\mathrm{X}-E[\mathrm{X}])^{\mathrm{T}}(\mathrm{X}-E[\mathrm{X}])]$ is the image covariance matrix, which is a n x n nonnegative definite matrix. Suppose that there are M training face images X_k $(k = 1, 2, \cdots, M)$, of which each face image is represented by an m x n matrix and $\overline{\mathrm{X}}$ denotes the average image by

$$\overline{\mathrm{X}} = \frac{1}{M}\sum_{k}\mathrm{X}_{\mathrm{k}}. \qquad (2)$$

Then G can be evaluated by

$$\mathrm{G} = \frac{1}{M}\sum_{k=1}^{M}(\mathrm{X}_k-\overline{\mathrm{X}})^{\mathrm{T}}(\mathrm{X}_k-\overline{\mathrm{X}}). \qquad (3)$$

It has been proven that the optimal value for the projection matrix B_{opt} is composed by the orthonormal eigenvectors $\mathrm{B}_1, \cdots, \mathrm{B}_d$ of G corresponding to the d largest eigenvalues [7]. Because the size of G is only n x n, computing its eigenvectors is very efficient. Also, like in PCA the value of d can be controlled by setting a threshold as follows:

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

where $\lambda_1, \lambda_2, \cdots, \lambda_n$ is the n largest eigenvalues of G and θ is a pre-set threshold [7].

B. $(2D)^2$PCA based on alternative 2DPCA

Let $\mathrm{X}_k = [(\mathrm{X}_k^{(1)})^{\mathrm{T}}(\mathrm{X}_k^{(2)})^{\mathrm{T}}\cdots(\mathrm{X}_k^{(m)})^{\mathrm{T}}]^{\mathrm{T}}$ and $\overline{\mathrm{X}} = [(\overline{\mathrm{X}}^{(1)})^{\mathrm{T}} (\overline{\mathrm{X}}^{(2)})^{\mathrm{T}}\cdots(\overline{\mathrm{X}}^{(m)})^{\mathrm{T}}]^{\mathrm{T}}$, where $\mathrm{X}_k^{(i)}$ and $\overline{\mathrm{X}}^{(i)}$ denote the i-th row vectors of X_k and $\overline{\mathrm{X}}$, respectively. Equation (3) can be rewritten as

$$\mathrm{G} = \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{m}(\mathrm{X}_k^{(i)}-\overline{\mathrm{X}}^{(i)})^{\mathrm{T}}(\mathrm{X}_k^{(i)}-\overline{\mathrm{X}}^{(i)}). \qquad (5)$$

Equation (5) reveals that the image covariance matrix G can be obtained from the product of row vectors of images, assuming the training images have zero mean, i.e. $\overline{\mathrm{X}} = (0)_{m\times n}$ [9]. Like this, the original 2DPCA is working in the row direction of images. So, a natural extension is to use the outer product between column vectors of images to construct G. Let $\mathrm{X}_k = [(\mathrm{X}_k^{(1)})(\mathrm{X}_k^{(2)})\cdots(\mathrm{X}_k^{(m)})]$ and $\overline{\mathrm{X}} = [(\overline{\mathrm{X}}^{(1)})(\overline{\mathrm{X}}^{(2)})\cdots (\overline{\mathrm{X}}^{(m)})]$, where $\mathrm{X}_k^{(i)}$ and $\overline{\mathrm{X}}^{(i)}$ denote the j-th column vectors of X_k and $\overline{\mathrm{X}}$, respectively. Then an alternative definition for image covariance matrix G is :

$$\mathrm{G} = \frac{1}{M}\sum_{k=1}^{M}\sum_{j=1}^{n}(\mathrm{X}_k^{(j)}-\overline{\mathrm{X}}^{(j)})(\mathrm{X}_k^{(j)}-\overline{\mathrm{X}}^{(j)})^{\mathrm{T}}. \qquad (6)$$

Now we will show how Eq. (6) can be at a similar way as in 2DPCA . Let $\mathrm{C} \in R^{m\times q}$ be a matrix with orthonormal columns. Projection the random matrix X onto C yields a q x n matrix, $\mathrm{Z}=\mathrm{C}^{\mathrm{T}}\mathrm{X}$ [9]. Similarly as in Eq. (1), the following criterion is adopted to find the optimal projection matrix C:

$$\begin{aligned} J(\mathrm{C}) &= trace\{E[(\mathrm{Z}-E[\mathrm{Z}])(\mathrm{Z}-E[\mathrm{Z}])^{\mathrm{T}}]\} \\ &= trace\{E[(\mathrm{C}^{\mathrm{T}}\mathrm{X}-E[\mathrm{C}^{\mathrm{T}}\mathrm{X}])(\mathrm{C}^{\mathrm{T}}\mathrm{X}-E[\mathrm{C}^{\mathrm{T}}\mathrm{X}])^{\mathrm{T}}]\} \qquad (7) \\ &= trace\{\mathrm{C}^{\mathrm{T}}E[(\mathrm{X}-E[\mathrm{X}])(\mathrm{X}-E[\mathrm{X}])^{\mathrm{T}}]\mathrm{C}\}, \end{aligned}$$

Similarly, the optimal projection matrix C_{opt} can be obtained by computing the eigenvectors $\mathrm{C}_1, \cdots, \mathrm{C}_q$ of Eq. (6) corresponding to the q largest eigenvalues. The value of q can be controlled by setting a threshold as in Eq. (4). Because the eigenvectors of Eq. (6) only reflect the information between columns of images, we say that the alternative 2DPCA is working in the column direction of images [9].

Suppose we have obtained the projection matrices B and C, projecting the $m \times n$ image X onto B and C simultaneously, yielding a $q \times d$ matrix A

$$A = C^{T}XB. \tag{8}$$

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

$$\hat{X} = CAB^{T}. \tag{9}$$

When used for face recognition, the matrix A is also called the feature matrix. After projection each training image $X_k\ (k = 1, 2, \cdots, M)$ onto B and C, we obtain the training feature matrices $A_k\ (k = 1, 2, \cdots, M)$ [9].

III. ICA

The PCA de-correlates the input data by leading eigenvectors of a covariance matrix of training dataset and exploits only the second-order statistics and it is optimal for datasets with Gaussian distribution. However, the distribution of facial images is non-Gaussian. There exist many high-order statistical dependencies among pixels of facial images. These dependencies can be further removed by using ICA, which can be viewed as a generalization of PCA [15, 16].

Fig. 1 shows the procedure of ICA. Suppose X can be represented as the linear combination of the components which are statistically independent (or as independent as possible), then the noise-free model of ICA can be described as

$$X = AS \tag{10}$$

where S is the vector of ICs, A is an unknown mixing matrix. In general ICs and the mixing matrix A are unknown. ICA aims to find a de-mixing matrix W such as

$$U = WX = WAS. \tag{11}$$

It can estimate the S with possible permutation and rescaling.

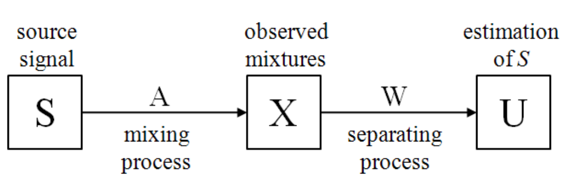

Figure 1. Blind source separation model.

To implement ICA, researchers have developed many algorithms based on two criteria: nonlinear de-correlation and maximum non-Gaussinanity [17-19]. Among them, the Fast-ICA which is based on maximum non-Gaussianity, has been dominantly used [18].

A. *Architectre I: Statistically Independent Basis Images*

Regardless of which algorithm is used to compute ICA, there are two different ways to apply ICA to face recognition. In architecture I, the input face images in X are considered to be a linear mixture of statistically independent basis images S combined by an unknown mixing matrix A. The ICA algorithm learns the weight matrix W, which is used to recover a set of independent basis images in the rows of U. In this architecture, the face images are variables and the pixel values provide observations for the variables. The source separation, therefore, is performed in face space. Projecting the input images onto the learned weight vectors produces the independent basis images. The compressed representation of a face image is a vector of coefficients used for linearly combining the independent basis images to generate the image. Bartlett and colleagues firstly apply PCA to project the data into a subspace of dimension *m* to control the number of independent components produced by ICA. The Fast-ICA algorithm is then applied to the eigenvectors to minimize the statistical dependence among the resulting basis images [18]. This use of PCA as a pre-processor in a two step process allows ICA to create subspaces of size *m* for any *m*. PCA de-correlates the input data; the remaining higher-order dependencies are separated by ICA [15].

B. *Architecture II: Statistically Independent Coefficients*

While the basis images obtained in architecture I are statistically independent, the coefficients that represent input images in the subspace defined by the basis images are not. The goal of ICA in architecture II is to find statistically independent coefficients for input data. In this architecture, the input is transposed from architecture I, that is, the pixels are variables and the images are observation. The source separation is performed on the pixels, and each row of the learned weight matrix W is an image. A, the inverse matrix of W, contains the basis images in its columns. The statistically independent source coefficients in S that comprise the input images are recovered in the columns of U. This architecture was used in [13] to find image filters that produced statistically independent outputs from natural scenes. In this work, ICA is performed on the PCA coefficients rather than directly on the input images to reduce the dimensionality [15].

IV. $(2D)^2$PCA-ICA

The proposed $(2D)^2$PCA-ICA is the method in which a face image matrix is de-correlated and the dimension is reduced. Then independent matrix is obtained by the Fast-ICA using the extracted final facial features [18]. A nearest neighbor classifier is used for recognition of each individual face in a facial features space. Fig. 2 shows the overall structure of the proposed model. Suppose training sample matrices represented by $X = [x_1, x_2, \cdots, x_M]$, and its corresponding independent components by $S = [s_1, s_2, \cdots, s_N]$.

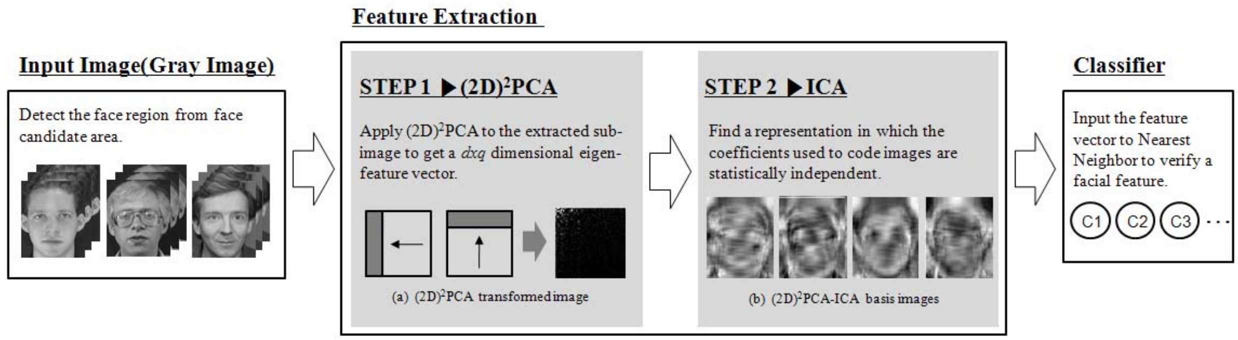

Figure 2. The Overall structure of the proposed $(2D)^2$PCA-ICA model

The relationship between S and X can be expressed as Eq. (10). From this relationship, each face image x_i can be represented as a linear combination of $s_1, s_2, \cdots, s_N$ with weighting $a_{i1}, a_{i2}, \cdots, a_{iN}$ respectively. Therefore, A is unknown M x N matrix with full rank, called the mixed-feature matrix. As mixing matrix A is an invertible matrix, we can get W by

$$\mathrm{A} = \mathrm{W}^{-1} \tag{12}$$

where W is called a transformation matrix.

In order to reduce the dimensionality of the input, instead of performing ICA directly on the p image pixels, ICA was performed on the PCA coefficients of the face images. We use the method of $(2D)^2$PCA to reduce the dimension of the training sample x_i, where $x_i \in R^{m \times n}, i = 1, 2, \cdots, M$. By way of Eq. (8), we have obtained q x d coefficient matrix F using the projection matrices B and C, projecting the m x n image X onto B and C simultaneously.

$$\mathrm{F} = \mathrm{C}^{\mathrm{T}}\mathrm{X}\mathrm{B} \tag{13}$$

Then, we use architecture with statistically independent ICA method sequentially. ICA attempts to make the outputs, U, as independent as possible. Hence, U is a factorial code for the face images. The representational code for test images is obtained by

$$\mathrm{U}_{\text{test}} = \mathrm{W}_{\mathrm{I}}\mathrm{X}_{\text{test}} \tag{14}$$

where X_{test} is zero-mean matrix of test images, and W_{I} is the weight matrix found by performing ICA on the training images. We choose the eigenvectors corresponding to the N largest eigenvalues to calculate the whitened matrix. That is

$$\mathrm{E} = \Lambda_{\mathrm{N}}^{-\frac{1}{2}} \mathrm{V}_{\mathrm{N}}^{\mathrm{T}} \tag{15}$$

where $\Lambda_{\mathrm{N}} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ 0 & & \ddots & \\ & & & \lambda_N \end{bmatrix}$, $\lambda_i (i = 1, 2, \cdots, N)$ denotes eigenvalues, and are arranged as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. If eigenvectors corresponding to the eigenvalues $\lambda_i (i = 1, 2, \cdots, N)$ are denoted by $v_i (i = 1, 2, \cdots, N)$, feature matrix comprising of eigenvectors can be shown as $\mathrm{V}_{\mathrm{N}} = [v_1, v_2, \cdots, v_N]$.

Finally, training samples are whitened as $E(x_i - \overline{x})^{\mathrm{T}}$, $i = 1, 2, \cdots, M$. And ICA is used to calculate mixing matrix A and the transformation W. Then we can obtain the principal independent matrix U as follows

$$\mathrm{U} = \mathrm{W}_{\mathrm{I}}\mathrm{E}. \tag{16}$$

Thus the optimal projection vectors of $(2D)^2$PCA-ICA are obtained by

$$\mathrm{Y} = [x_i - \overline{x}]\mathrm{U}^{\mathrm{T}}. \tag{17}$$

After transformation by $(2D)^2$PCA-ICA, a series of projected feature vectors are obtained for each image, and the matrix $\mathrm{Q} = [\mathrm{Y}_1, \mathrm{Y}_2, \cdots, \mathrm{Y}_N]$ is formed. Then a nearest neighbor classifier is used in classification.

Here, the distance between two arbitrary feature matrices is defined by

$$\mathrm{D}(\mathrm{B}_{\text{test}}, \mathrm{B}) = \sum_{k=1}^{N} \left\| \mathrm{Y}_{ktest} - \mathrm{Y}_k \right\|. \tag{18}$$

Suppose the class of a training sample represented by w_k.

V. EXPERIMENTS AND ANALYSIS

In PCA, 2-D face image matrix must be transformed into 1-D image vector. The resulting image vectors of faces usually lead to a high-dimensional image vector space, where it is difficult to evaluate the whitened matrix. It does not imply that the whitened matrices can be accurately evaluated in this way since the whitened matrices are statistically determined by the eigenvectors. And the method of $(2D)^2$PCA is more suitable for small sample size problems in face recognition. But it is more difficult to reduce dimension further by way of $(2D)^2$PCA method. And it cannot get rid of the influence of illumination.

Fortunately, the technique of the proposed $(2D)^2$PCA-ICA can obtain the whitened matrix through the covariance matrix using $(2D)^2$PCA, and the size of the covariance matrix is smaller than PCA. Because ORL and Yale B face database contain the images under various facial expression and lighting conditions, the method presented is suitable for feature extraction and recognition.

A. Experiments on ORL Database

The proposed $(2D)^2$PCA-ICA method is used for face recognition and tested on ORL face image database (http://www.cl.cam.ac.uk), which is used to evaluate the performance of the proposed method. The ORL database contains images from 40 individuals, each providing 10 different images. All images are grayscale and normalized to a resolution of 112x92 pixels. Five images of one person in ORL are shown in Fig. 3 (a). In an experiment, the first five images samples per class are used for training data, and the remaining images for test data. Thus, the total number of training samples and testing samples are both 200. We compare the proposed method with another feature extraction method. Finally a nearest neighbor classifier is employed for face recognition in a facial feature space.

Table 1 gives the performance comparisons of five approaches on recognition rate for several cases using different dimensions of feature vector. In this experiment, the number of projection vectors in all approaches is controlled by the value of θ (in Eq. (4)), which is set of 0.95. The performance of the $(2D)^2$PCA-ICA is better than another methods. And the results are plotted in Fig. 4 obtained from classification experiments under different dimensions for $(2D)^2$PCA, PCA-ICA and $(2D)^2$PCA-ICA, which shows the relationship between the recognition rate and dimensions. As shown in Fig. 4, $(2D)^2$PCA-ICA obtains better accuracy than both $(2D)^2$PCA and PCA-ICA under the same dimensions of feature vectors.

B. Experiments on Yale B Database


(a)
(b)

Figure 3. Samples from face database (a) ORL (b) Yale B.

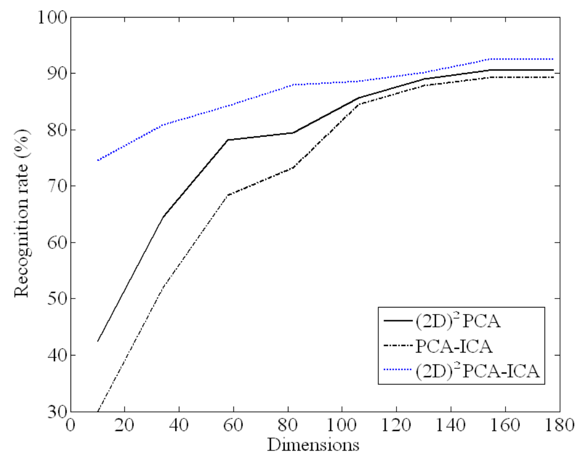

Figure 4. Comparisons of recognition rate between $(2D)^2$PCA, PCA-ICA and $(2D)^2$PCA-ICA under different dimensions.

Another experiment is performed using the Yale B face database as shown in Fig. 3 (b), which contains 200 images of 10 individuals (each person has 20 different images) under various lighting conditions. Each image is manually cropped and resized to 60x50 pixels in this experiment. In this experiment, we also use the first five image samples per class for training and the remaining images for testing. Table 1 gives the performance comparisons of five approaches on recognition rate according to different dimension of feature vector.

TABLE I. COMPARISONS OF FIVE APPROACHES ON ORL AND YALE B DATABASE

Approach	**ORL face Database**		**Yale B face Database**	
	success rate (%)	***dimension of feature vector***	***success rate (%)***	***dimension of feature vector***
PCA	88.0	110	86.0	40
ICA	86.0	98	87.0	48
PCA-ICA	89.2	98	90.0	40
$(2D)^2$PCA	90.5	27x24	89.0	32x27
$(2D)^2$PCA-ICA	92.5	26x18	91.0	30x21

VI. CONCLUSION

In this paper, an efficient face recognition approach called $(2D)^2$PCA-ICA is proposed. In the proposed face recognition model, $(2D)^2$PCA-ICA properly extracts facial features, where $(2D)^2$PCA is simpler and more straightforward in dimension reduction based on 2-D image than the convention 1-D PCA. Moreover, ICA can sequentially remove the high-order statistical dependencies to produce a sparse and independent code for subsequent pattern discrimination. Furthermore the computational efficiency of the proposed $(2D)^2$PCA-ICA is remarkably improved compared with (1-D)PCA. As well, the proposed model shows better face recognition performance than other methods based on PCA and ICA. In the future work, we are considering an incremental scheme to properly deal with a large-scale database, which can incrementally learn high-dimensional data without computing the corresponding covariance matrix and without knowing a priori knowledge about the data in advance.

REFERENCES

[1] I. T. Jolliffe, "Principal Component Analysis," Springer NewYork, 1986.

[2] M. Kirby, L. Sirovich, "Application of the KL procedure for the characterization of human faces," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 1, pp. 103–108, 1990.

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

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

[5] I. Dagher and R. Nachar, "Face Recognition Using IPCA-ICA Algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 6, pp. 996-1000, Jun 2006.

[6] C. Liu, "Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 5, pp. 572-581, May 2004.

[7] J. Yang, D. Zhang, A. F. Frangi, J. Y. Yang, "Two-dimensional PCA: a new approach to appearancebased face representation and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 1, pp. 131–137, 2004.

[8] D. Q. Zhang, S. C. Chen, J. Liu, "Representing image matrices: eigenimages vs. eigenvectors," Proceedings of the Second International Symposium on Neural Networks (ISNN'05), vol. 2, Chongqing, China, pp. 659–664, 2005.

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

[10] H. B. Barlow, "Unsupervised learning," Neural Comput., vol. 1, pp. 295–311, 1989.

[11] J. J. Atick and A. N. Redlich, "What does the retina know about natural scenes?," Neural Comput., vol. 4, pp. 196–210, 1992.

[12] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," Neural Comput., vol. 7, no. 6, pp. 1129–1159, 1995.

[13] A. J. Bell and T. J. Sejnowski, "The independent components of natural scenes are edge filters," Vision Res., vol. 37, no. 23, pp. 3327–3338, 1997.

[14] J. -F. Cardoso, "Blind Signal Separation, Statistical Principles," Proc. IEEE, vol. 9, no. 10, 1998.

[15] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by Independent Component Analysis," IEEE Trans. Neural Netw., vol. 12, pp. 1450-1464, 2002.

[16] J. -F. Cardoso, "Infomax and Maximum Likelihood for Source Separation," IEEE Signal Proc. Let., vol. 4, pp. 112-114, 1997.

[17] A. Hyvärinen, "Survey on independent component analysis," Neural Comput. Surveys2, pp. 94–128, 1999.

[18] A. Hyvärinen and E. Oja, " A fast fixed-point algorithm for independent component analysis," Neural Comput., vol. 9, no. 7, pp. 1483-1492, 1997.

[19] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," Neural Netw., vol. 13, pp. 411-430, 2000.