2007 IEEE International Symposium
on Signal Processing and Information Technology

Kernel-based 2DPCA for Face Recognition

Vo Dinh Minh Nhat
Dept. of Computer Engineering
Kyung Hee University - Suwon, Korea
vdmnhat@oslab.khu.ac.kr

SungYoung Lee
Dept. of Computer Engineering
Kyung Hee University - Suwon, Korea
sylee@oslab.khu.ac.kr

Abstract *— Recently, in the field of face recognition, Two-dimensional Principal Component Analysis (2DPCA) has been proposed in which image covariance matrices can be constructed directly using original image matrix. In contrast to the covariance matrix of traditional PCA, the size of the image covariance matrix using 2DPCA is much smaller. As a result, it is easier to evaluate the covariance matrix accurately, computation cost is reduced and the performance is also improved. In an effort to improve and perfect the performance of face recognition system, in this paper, we propose a Kernel-based 2DPCA (K2DPCA) method which can extract nonlinear principal components based directly on input image matrices. Similar to Kernel PCA, K2DPCA can extract nonlinear features efficiently instead of carrying out the nonlinear mapping explicitly. Experiment results show that our method achieves better performance in comparison with the other approaches.*

Keywords *— PCA, Kernel PCA, 2DPCA, Face Recognition.*

I. Introduction

Principal component analysis (PCA), also known as Karhunen-Loeve expansion, is a classical feature extraction and data representation technique widely used in the areas of pattern recognition and computer vision. Sirovich and Kirby [1][2] first used PCA to efficiently represent pictures of human faces. They argued that any face image could be reconstructed approximately as a weighted sum of a small collection of images that define a facial basis (eigenimages), and a mean image of the face. Within this context, Turk and Pentland [3] presented the well-known Eigenfaces method for face recognition in 1991. Since then, PCA has been widely investigated and has become one of the most successful approaches in face recognition [4][5][6][7]. Penev and Sirovich [8] discussed the problem of the dimensionality of the "face space" when eigenfaces are used for representation. Zhao and Yang [9] tried to account for the arbitrary effects of illumination in PCA-based vision systems by generating an analytically closedform formula of the covariance matrix for the case with a special lighting condition and then generalizing to an arbitrary illumination via an illumination equation. However, Wiskott et al. [10] pointed out that PCA could not capture even the simplest invariance unless this information is explicitly provided in the training data. They proposed a technique known as elastic bunch graph matching to overcome the weaknesses of PCA. Recently, two PCA-related methods, independent component analysis (ICA) and kernel principal component analysis (Kernel PCA) have been of wide concern. Bartlett et al. [11] and Draper et al. [12] proposed using ICA for face representation and found that it was better than PCA when cosines were used as the similarity measure (however, their performance was not significantly different if the Euclidean distance is used). Yang [13] used Kernel PCA for face feature extraction and recognition and showed that the Kernel Eigenfaces method outperforms the classical Eigenfaces method. However, ICA and Kernel PCA are both computationally more expensive than PCA. The experimental results in [13] showed the ratio of the computation time required by ICA, Kernel PCA, and PCA is, on average, 8.7: 3.2: 1.0.

In all previous PCA-based face recognition technique, the 2D face image matrices must be previously transformed into 1D image vectors. The resulting image vectors of faces usually lead to a high dimensional image 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. Fortunately, the eigenvectors can be calculated efficiently using the SVD techniques and the process of generating the covariance matrix is actually avoided. However, this does not imply that the eigenvectors can be evaluated accurately in this way since the eigenvectors are statistically determined by the covariance matrix, no matter what method is adopted for obtaining them. So recently in [14], a new PCA approach called 2DPCA, is developed for image feature extraction. As opposed to conventional PCA, 2DPCA is based on 2D matrices rather than 1D vectors. That is, the image matrix does not need to be transformed into vector. Instead, an image covariance matrix can be constructed directly using original image matrices. In contrast to the covariance matrix of PCA, the size of the image covariance matrix using 2DPCA is much smaller. As a result, 2DPCA has two important advantages over PCA. First, it is easier to evaluate the covariance matrix accurately. Second, less time is required to determine the corresponding eigenvectors. In an effort to improve and perfect the performance of face recognition system, in this paper, we propose a Kernel-based 2DPCA (K2DPCA) method which can extract nonlinear principal components based directly on input image matrices. Similar to Kernel PCA, K2DPCA can extract nonlinear features efficiently instead of carrying out the nonlinear mapping explicitly. The remainder of this paper is organized as follows: In Section 2, the traditional PCA and KPCA methods are reviewed. Section 3 comes with 2DPCA.

978-1-4244-1835-0/07/$25.00 ©2007 IEEE

The proposed Kernel-based 2DPCA method is described in Section 4. In Section 5, experimental results are presented for the ORL and Yale face image databases to demonstrate the effectiveness of our methods. Finally, conclusions are presented in Section 6.

II. PCA AND KERNEL PCA

One approach to cope with the problem of excessive dimensionality of the image space is to reduce the dimensionality by combining features. Linear combinations are particular, attractive because they are simple to compute and analytically tractable. In effect, linear methods project the high-dimensional data onto a lower dimensional subspace. Suppose that we have Nsample images $\{x_1, x_2, ..., x_N\}$ taking values in an n-dimensional image space. Let us also consider a linear transformation mapping the original n-dimensional image space into an m-dimensional feature space, where $m < n$. The new feature vectors $y_k \in \mathrm{R}^m$ are defined by the following linear transformation:

$$y_k = W^T x_k \tag{1}$$

where $k = 1, 2, ..., N$ and $W \in \mathrm{R}^{nxm}$ is a matrix with orthonormal columns. Different objective functions will yield different algorithms with different properties. PCA aims to extract a subspace in which the variance is maximized. Its objective function is as follows:

$$W_{opt} = [w_1 w_2 ... w_m] = \arg\max_W \left| W^T S_t W \right| \tag{2}$$

with the total scatter matrix is defined as

$$S_t = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \tag{3}$$

and $\mu \in \mathrm{R}^n$ is the mean of all samples. The optimal projection $W_{opt} = [w_1 w_2 ... w_m]$ is the set of n-dimensional eigenvectors of S_t corresponding to the m largest eigenvalues, i.e:.

$$S_t w_i = \lambda_i w_i \quad i = 1, 2, ..., m \tag{4}$$

In kernel PCA, each vector x_i is projected from the input space, $\mathcal{T}$or $\Re^n$, to a high dimensional feature space, $\mathcal{F}$or $\Re^f$, by a nonlinear mapping function $\Phi : \Re^n \to \Re^f$, $f > n$. In $\Re^f$, the corresponding eigenvalue problem (4) becomes

$$S_t^\Phi w^\Phi = \lambda w^\Phi \tag{5}$$

Without losing the generality, we assume that the projected samples $\Phi(x_i)$ are centered in $\Re^f$ (see [15] for a method to center the vector $\Phi(x_i)$ in $\Re^f$). The total scatter matrix can be re-calculated in $\Re^f$ as follow:

$$S_t^\Phi = \sum_{k=1}^{N} \Phi(x_k)\Phi(x_k)^T = A_t^\Phi \left(A_t^\Phi\right)^T \tag{6}$$

where $A_t^\Phi = [\Phi(x_1), .., \Phi(x_N)]$ is a matrix whose columns are $\Phi(x_i)$. We can see that, all solution w^Φ of (5) with $\lambda \neq 0$ lie in the span of $\Phi(x_1), .., \Phi(x_N)$, and there exist coefficient vector $\alpha = [\alpha_i, .., \alpha_N]^T$ such that

$$w^\Phi = \sum_{i=1}^{N} \alpha_i \Phi(x_i) = A_t^\Phi \alpha \tag{7}$$

Denoting an NxN matrix K by

$$K_{ij} = k(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) = \left(A_t^\Phi\right)^T A_t^\Phi \tag{8}$$

Then the kernel PCA problem (5) becomes

$$\begin{aligned} & A_t^\Phi \left(A_t^\Phi\right)^T A_t^\Phi \alpha = \lambda A_t^\Phi \alpha \\ \Leftrightarrow\ & A_t^\Phi K \alpha = A_t^\Phi \lambda \alpha \\ \Leftarrow\ & K\alpha = \lambda\alpha \end{aligned} \tag{9}$$

So firstly, we solve the equation $K\alpha = \lambda\alpha$. We can now projected the vectors in $\Re^f$ to a lower dimensional space spanned by the eigenvectors w^Φ. Let x be a test sample whose projection is $\Phi(x)$ in $\Re^f$, then the projection of $\Phi(x)$ onto the eigenvectors w^Φ is:

$$w^\Phi \Phi(x) = \sum_{i=1}^{N} \alpha_i \left(\Phi(x_i)^T \Phi(x)\right) = \sum_{i=1}^{N} \alpha_i k(x_i, x_j) \tag{10}$$

We can extract the first m $(1 \leq m \leq N-1)$ nonlinear principal components corresponding to first m non-increasing eigenvalues of (5) using the kernel function without the expensive operation that explicitly projects samples to high dimensional space $\Re^f$. Some kernel functions can be seen in Table 1., and in this paper, we use polynomial kernel function. Compared to other techniques for nonlinear feature extraction,

Kernel	$k(x_1, x_2)$
Polynomial	$(x_1^T x_2)^d$
Gaussian (radial basis function)	$e^{-\frac{\|x_1 - x_2\|^2}{2\sigma^2}}$
Sigmoid	$\tanh(ax_1^T x_2 + b)$

Table 1. Some kernel functions can be used in Kernel PCA.

kernel PCA has the advantages that (1) it does not require nonlinear optimization but just the solution of an Eigenvalue problem and (2) by the possibility to use different kernels it comprises a fairly general class of nonlinearities that can be used [15].

III. TWO-DIMENSIONAL PCA

In 2D approach, the image matrix does not need to be previously transformed into a vector, so a set of N sample images is represented as $\{X_1, X_2, ..., X_N\}$ with $X_i \in \mathrm{R}^{kxs}$, which is a matrix space of size kxs. The total scatter matrix is re-defined as

$$G_t = \sum_{i=1}^{N} (X_i - \mu_X)^T (X_i - \mu_X) \tag{11}$$

with $\mu_X = \frac{1}{N} \sum_{i=1}^{N} X_i \in \mathrm{R}^{kxs}$ is the mean image of all samples. $G_t \in \mathrm{R}^{sxs}$ is also called image covariance (scatter) matrix. A

Fig. 1. Ten sample images from Yale face database

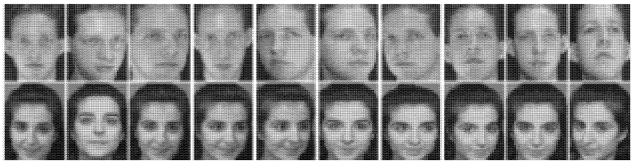

Fig. 2. Twenty sample images from ORL face database

linear transformation mapping the original kxs image space into an kxm feature space, where $m < s$. The new feature matrices $Y_i \in \mathrm{R}^{kxm}$ are defined by the following linear transformation:

$$Y_i = (X_i - \mu_X)W \in \mathrm{R}^{kxm} \tag{12}$$

where $i = 1, 2, ..., N$ and $W \in \mathrm{R}^{sxm}$ is a matrix with orthonormal columns. In 2DPCA, the projection W_{opt} is chosen to maximize $tr(W^T G_t W)$. The optimal projection $W_{opt} = [w_1 w_2 ... w_m]$ with $\{w_i \,|i = 1, 2, ..., m\}$ is the set of s-dimensional eigenvectors of G_t corresponding to the $m$largest eigenvalues. After a transformation by 2DPCA, a feature matrix is obtained for each image. Then, a nearest neighbor classifier is used for classification. Here, the distance between two arbitrary feature matrices Y_i and Y_j is defined by using Euclidean distance as follows:

$$d(Y_i, Y_j) = \sqrt{\sum_{u=1}^{k}\sum_{v=1}^{s}(Y_i(u,v) - Y_j(u,v))^2} \tag{13}$$

Given a test sample Y_t, if $d(Y_t, Y_c) = \min_j d(Y_t, Y_j)$, then the resulting decision is Y_t belongs to the same class as Y_c.

IV. Kernel-based 2DPCA

Kernel-based 2DPCA proposed here is an unsupervised feature extraction method closely related to 2DPCA. And it is based on the following ideas:

- Nonlinearly mapping input matrix space $\mathcal{T}$ or $\Re^{kxs}$ to a high-dimensional matrix feature space $\mathcal{F}$ or $\Re^{kxf}$, $\Psi : \mathcal{T} \to \mathcal{F}$, where standard 2DPCA is performed; this requires dot products.
- Computing dot products in high-dimensional matrix feature space $\mathcal{F}$ via a matrix-based kernel function in input matrix space $\mathcal{T}$,$k : \mathcal{T}x\mathcal{T} \to \Re^{kxk}$. This allows to perform 2DPCA efficiently on the data set $\{\Psi(X_i)\}_{i=1}^N$ using the dot-product but never explicitly calculating Ψ for any input matrix. The nonlinear nature of map Ψ means that the associated component analysis back in input space $\mathcal{T}$ is nonlinear.

Let us analyze more in detail the procedure. In kernel-based 2DPCA, each image X_i is projected from the input matrix space, $\mathcal{T}$or $\Re^{kxs}$, to a high dimensional feature space, $\mathcal{F}$or $\Re^{kxf}$, by a nonlinear mapping function $\Psi : \Re^{kxs} \to \Re^{kxf}$, $f > s$. Let $X_i^{j*} \in \Re^{1xs}$ be the j^{th} row of matrix $X_i \in \Re^{kxs}$, where $j = 1..k$, then nonlinear mapping function $\Psi : \Re^{kxs} \to \Re^{kxf}$ can be defined as follow:

$$\Psi(X_i) = \begin{bmatrix} \Phi\left((X_i^{1*})^T\right)^T \\ ... \\ \Phi\left((X_i^{k*})^T\right)^T \end{bmatrix} \in \Re^{kxf} \tag{14}$$

Note that here, $\Phi : \Re^s \to \Re^f$ is a nonlinear mapping function defined in a similar way as in section 2. With a similar way used in KPCA, the total scatter matrix in K2DPCA can be re-calculated in $\Re^{kxf}$ as follow:

$$\begin{aligned} G_t^\Phi &= \sum_{i=1}^{N} \Psi(X_i)^T \Psi(X_i) \\ &= \sum_{i=1}^{N} \begin{bmatrix} \Phi\left((X_i^{1*})^T\right)^T \\ ... \\ \Phi\left((X_i^{k*})^T\right)^T \end{bmatrix}^T \begin{bmatrix} \Phi\left((X_i^{1*})^T\right)^T \\ ... \\ \Phi\left((X_i^{k*})^T\right)^T \end{bmatrix} \\ &= \sum_{i=1}^{N} \left[\Phi\left((X_i^{1*})^T\right), ..., \Phi\left((X_i^{k*})^T\right)\right] \begin{bmatrix} \Phi\left((X_i^{1*})^T\right)^T \\ ... \\ \Phi\left((X_i^{k*})^T\right)^T \end{bmatrix} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{k} \Phi\left((X_i^{j*})^T\right) \Phi\left((X_i^{j*})^T\right)^T \end{aligned} \tag{15}$$

From above equation (15), because $(X_i^{k*})^T$ is vector in $\Re^s$, we can see that the total scatter matrix G_t^Φ can be calculated on the rows of all the training images. By this observation, K2DPCA could be performed based on KPCA. Consider each row of all training image matrices as a column vector sample, K2DPCA can be done by performing KPCA on these reformed samples. A summary of K2DPCA algorithm can be seen as follow:

Idea - Consider each row of all training image matrices as a column-vector sample and apply KPCA.

Input - A set of Nsample images is represented as $\{X_1, X_2, ..., X_N\}$ with $X_i \in \mathrm{R}^{kxs}$. Let $r_l = \left(X_i^{j*}\right)^T \in \Re^s$, where $(i = 1..N, j = 1..k)$ and $l = k(i-1)+j$, be a column vector which is the transpose of the row j^{th} of image matrix i^{th}.

Algorithm

- Centering projected samples $\Phi(r_l)$ (see [15] for a method to center the vector $\Phi(r_l)$)
- Define kernel matrix $K \in \Re^{kNxkN}$ by

$$\begin{aligned} K_{ij} &= k(r_i, r_j) = \Phi(r_i)^T \Phi(r_j) \\ i, j &= 1..kN \quad and \quad \Phi : \Re^s \to \Re^f \end{aligned} \tag{16}$$

- Solve the eigen-problem $K\alpha = \lambda\alpha$. Remember that $\alpha \in \Re^{kN}$ and $w^{\Phi} = \sum_{i=1}^{kN} \alpha_i \Phi(r_i) \in \Re^{f}$.
- Projecting the image in $\Re^{kxf}$ to a lower dimensional space spanned by the eigenvectors w^{Φ}. Let $X \in \Re^{kxs}$ be a sample whose projection is $\Psi(X)$ in $\Re^{kxf}$, then the projection of $\Psi(X)$ onto the eigenvectors w^{Φ} is:

$$\begin{aligned}\Psi(X_i)w^{\Phi} &= \begin{bmatrix} \Phi\left((X_i^{1*})^T\right)^T \\ \dots \\ \Phi\left((X_i^{k*})^T\right)^T \end{bmatrix} w^{\Phi} \\ &= \begin{bmatrix} \Phi\left((X_i^{1*})^T\right)^T w^{\Phi} \\ \dots \\ \Phi\left((X_i^{k*})^T\right)^T w^{\Phi} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{t=1}^{nN} \alpha_t \left(\Phi\left((X_i^{1*})^T\right)^T \Phi(r_t)\right) \\ \dots \\ \sum_{t=1}^{nN} \alpha_t \left(\Phi\left((X_i^{k*})^T\right)^T \Phi(r_t)\right) \end{bmatrix} \in \Re^k \end{aligned} \quad (17)$$

V. Experimental results

This section evaluates the performance of PCA algorithm [1], KPCA algorithm [13], 2DPCA [14], our new approach K2DPCA based on using Yale face database and ORL face database. The size of each cropped image in both ORL and Yale database is $32x32$ pixels, with 256 gray levels per pixel. In this paper, we apply the nearest-neighbor classifier for its simplicity. The Euclidean metric is used as our distance measure. In short, the recognition process has three steps. First, we calculate the face subspace from the training set of face images; then the new face image to be identified is projected into m-dimensional subspace. Finally, the new face image is identified by a nearest neighbor classifier. Some sample images from Yale and ORL databases are shown in Fig. 1. and Fig. 2.

A. Yale Face Database

The Yale face Database contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink. A random subset with $k(k = 2, 3, 4, 5)$ images per individual was taken with labels to form the training set. The rest of the database was considered to be the testing set. 10 times of random selection for training examples were performed and the average recognition result was recorded. The training samples were used to learn the subspace. The testing samples were then projected into the low-dimensional representation subspace. We tested the recognition rates with different number of training samples and show the best results obtained by PCA, KPCA, 2DPCA, and our approach K2DPCA in Table 2.

Table 2. Comparison of the top recognition accuracy (%) on Yale database.

k	2	3	4	5
PCA	76.30	83.33	84.76	87.78
KPCA	76.30	83.33	85.15	88.19
2DPCA	76.30	83.33	88.57	88.89
K2DPCA	76.30	85.12	89.00	89.95

B. ORL Face Database

In the ORL database, there are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). A random subset with $k(k = 2, 3, 4, 5)$ images per individual was taken with labels to form the training set. The rest of the database was considered to be the testing set. 10 times of random selection for training example were performed and the average recognition result was recorded. The experimental protocol is the same as before. The best recognition result of each method are shown in Table 3.

Table 3. Comparison of the top recognition accuracy (%) on ORL database.

k	2	3	4	5
PCA	72.81	76.43	85.83	89.50
KPCA	72.81	77.56	86.12	89.52
2DPCA	72.19	76.43	86.25	89.00
K2DPCA	74.18	78.71	88.83	91.13

VI. Conclusions

In this paper, we propose a Kernel-based 2DPCA (K2DPCA) method which can extract nonlinear principal components based directly on input image matrices. Similar to Kernel PCA, K2DPCA can extract nonlinear features efficiently instead of carrying out the nonlinear mapping explicitly. We also proved that K2DPCA could be implemented by using KPCA technique. The experiments based on ORL and Yale face databases show clear improvements of recognition systems. In the future, we will study kernel approach for two-dimensional Linear Discriminant Analysis.

Acknowledgments

This research was supported by the MIC(Ministry of Information and Communication), Korea, Under the ITFSIP (IT Foreign Specialist Inviting Program) Supervised by the IITA (Institute of Information Technology Advancement).

References

[1] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of Optical Soc. Am*, vol. 4, pp. 519–524, 1987.

[2] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103–108, 1990.
[3] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
[4] A. Pentland, "Looking at people: Sensing for ubiquitous and wearable computing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 107–119, 2000.
[5] M. A. Grudin, "On internal representations in face recognition systems," *Pattern Recognition*, vol. 33, no. 7, pp. 1161–1177, 2000.
[6] G. W. Cottrell and M. K. Fleming, "Face recognition using unsupervised feature extraction," in *Proceedings of the International Neural Network Conference, Paris*. Kluwer, 1990, pp. 322–325.
[7] D. Valentin, H. Abdi, A. O'Toole, and G. Cottrell, "Connectionist models of face processing: A survey," vol. 27, no. 9, pp. 1209–1230, September 1994.
[8] P. S. Penev and L. Sirovich, "The global dimensionality of face space," in *FG '00: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*. Washington, DC, USA: IEEE Computer Society, 2000, p. 264.
[9] L. Zhao and Y. Yang, "Theoretical analysis of illumination in pca-based vision systems," vol. 32, no. 4, pp. 547–564, April 1999.
[10] L. Wiskott, J. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," vol. 19, no. 7, pp. 775–779, July 1997.
[11] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, pp. 1450–1464, 2002. [Online]. Available: http://mplab.ucsd.edu/projects-home/project1/publications/pdfs/c-BartlettMovellanSejnowski2002-4.pdf
[12] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, "Recognizing faces with pca and ica," *Comput. Vis. Image Underst.*, vol. 91, no. 1-2, pp. 115–137, 2003.
[13] M.-H. Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 2002, p. 215.
[14] J. Yang, D. Zhang, A. F. Frangi, and J. yu Yang, "Two-dimensional pca: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, 2004.
[15] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.