Two-dimensional Weighted PCA algorithm for Face Recognition

Vo Dinh Minh Nhat
Dept. of Computer Engineering
Kyung Hee University
Giheung-Eup, Yongin-Si, Gyeonggi-Do, 449-701, Korea
vdmnhat@oslab.khu.ac.kr

SungYoung Lee
Dept. of Computer Engineering
Kyung Hee University
Giheung-Eup, Yongin-Si, Gyeonggi-Do, 449-701, Korea
sylee@oslab.khu.ac.kr

Abstract **- Principle Component Analysis (PCA) technique is an important and well-developed area of image recognition and to date many linear discrimination methods have been put forward. Basically, in PCA the image always needs to be transformed into 1D vector, however recently two-dimensional PCA (2DPCA) technique have been proposed. In 2DPCA, PCA technique is applied directly on the original images without transforming into 1D vector. In this paper, we propose a new 2DPCA-based method that can improve the performance of the 2DPCA approach. In face recognition where the training data are labeled, a projection is often required to emphasize the discrimination between the clusters. Both PCA and 2DPCA may fail to accomplish this, no matter how easy the task is, as they are unsupervised techniques. The directions that maximize the scatter of the data might not be as adequate to discriminate between clusters. So we proposed a new 2DPCA-based scheme which can straightforwardly take into consideration data labeling, and makes the performance of recognition system better. Experiment results show our method achieves better performance in comparison with the 2DPCA approach with the complexity nearly as same as that of 2DPCA method.**

Index Terms - Principle component analysis (PCA), Two-dimensional PCA (2DPCA), Two-dimensional Weighted PCA, 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 [14] 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 [14] 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 [16], 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.

This research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment)

Corresponding Authors : Vo Dinh Minh Nhat (vo_dinhminhnhat@yahoo.com), and SungYoung Lee (sylee@oslab.khu.ac.kr)

However, in face recognition where the data are labeled, a projection is often required to emphasize the discrimination between the clusters. Both PCA and 2DPCA may fail to accomplish this, no matter how easy the task is, as they are unsupervised techniques. The directions that maximize the scatter of the data might not be as adequate to discriminate between clusters. In this paper, our proposed approach can straightforwardly take into consideration data labeling, which makes the performance of recognition system better. The remainder of this paper is organized as follows: In Section 2, the 2DPCA method is reviewed. The idea of the proposed method and its algorithm are described in Section 3. In Section 4, experimental results are presented on the ORL and Yale face databases to demonstrate the effectiveness of our method. Finally, conclusions are presented in Section 5.

II. Two-dimensional PCA

In this section, we review the basic notions, essential mathematical background and algorithm of 2DPCA approach that is needed for subsequent derivations in next sections.

Theorem 1. *Let A be an nxn symmetric matrix. Denoted by $\lambda_1 \geq \ldots \geq \lambda_n$ its sorted eigenvalues, and by $w_1, \ldots, w_n$ the corresponding eigenvectors. Then $w_1, \ldots, w_m (m < n)$ are the maximizer of the constrained maximization problem $\max tr(W^T A W)$ subject to $W^T W = I$.*

For the proof, we can reference [18].

In 2DPCA 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 $X_i \in \mathbb{R}^{kxs}$. The total scatter matrix is defined as

$$G_T = \sum_{i=1}^{N} (X_i - \mu_X)(X_i - \mu_X)^T \qquad (1)$$

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

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

$$Y_i = W^T (X_i - \mu_X) \in \mathbb{R}^{mxs} \qquad (2)$$

where $i = 1, 2, ..., N$ and $W \in \mathbb{R}^{kxm}$ is a matrix with orthonormal columns. In 2DPCA, the projection W_{opt} is chosen to maximize $tr(W^T G_T W)$. By Theorem 1, we have $W_{opt} = [w_1 w_2 ... w_m]$ with $\{w_i \,|\, i = 1, 2, ..., m\}$ is the set of *n*-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 :

$$d(Y_i, Y_j) = \sqrt{\sum_{u=1}^{k} \sum_{v=1}^{s} (Y_i(u,v) - Y_j(u,v))^2} \qquad (3)$$

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.

III. Two-dimentional Weighted PCA

In the following part, we present our proposed method. Firstly we will take a look at some necessary background. Let $A, B \in \mathbb{R}^{mxn}$, then A_{ci} and A_{rj} are i^{th} column vector and j^{th} row vector of matrix A. The Euclidean distance between A and B is defined as follows :

$$d(A, B)^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} (A_{ij} - B_{ij})^2 \qquad (4)$$

The *Laplacian* is a key entity for describing pairwise relationships between data elements. This is a symmetric positive-semidefinite matrix, characterized by having zero row and column sums.

Lemma 1. *Let L be an nxn Laplacian, and let* $B \in \mathbb{R}^{mxn}$. Then we have the following equation :

$$tr(BLB^T) = \sum_{i<j} -L_{ij} d(B_{ci}, B_{cj})^2 \qquad (5)$$

Proof. Let $z = [z_1 z_2 ... z_n]^T \in \mathbb{R}^n$ then we have

$$\begin{aligned} z^T L z &= \sum_i L_{ii} z_i^2 + 2 \sum_{i<j} L_{ij} z_i z_j \\ &= \sum_{i<j} -L_{ij} (z_i^2 + z_j^2) + 2 \sum_{i<j} L_{ij} z_i z_j \\ &= \sum_{i<j} -L_{ij} (z_i - z_j)^2 \end{aligned} \qquad (6)$$

By applying (5) we have

$$\begin{aligned} tr(BLB^T) &= \sum_{k=1}^{m} B_{rk} L B_{rk}^T \\ &= \sum_{i<j} \sum_{k=1}^{m} -L_{ij} (B_{ki} - B_{kj})^2 \\ &= \sum_{i<j} -L_{ij} d(B_{ci}, B_{cj})^2 \end{aligned} \qquad (7)$$

Proof is done.

Lemma 2 . We define a NxN unit Laplacian, denoted by L^u, as $L^u = N\delta_{ij} - 1$, with δ_{ij} is the Kronecker delta (defined as 1 for $i = j$ and as 0 otherwise), and $A \in \mathbb{R}^{nxN}$ with zero mean column (i.e. sum of all column vectors is a zero vector). We have

$$AL^u A^T = A(NI_N - U)A^T = NS_T - AUA^T = NS_T \qquad \textbf{(8)}$$

with I_N is identity matrix and U is a matrix of all ones. The last equality is due to the fact that the coordinates are centered. Proof is clear.

Let define A_i as follows :

$A_i = [((X_1)_{ci} - (\mu_X)_{ci}) \ldots \quad ((X_N)_{ci} - (\mu_X)_{ci})] \in \mathbb{R}^{kxN}$, and B_i be a matrix which is formed by all the column i^{th} of each matrix Y_i

$$B_i = [(Y_1)_{ci} \ldots \quad (Y_N)_{ci}] \in \mathbb{R}^{mxN} \qquad \textbf{(9)}$$

The image scatter matrix G_T could be re-written as follow :

$$\begin{aligned} G_T &= \sum_{i=1}^{N}(X_i - \mu_X)(X_i - \mu_X)^T \\ &= \sum_{i=1}^{N}\sum_{j=1}^{s}[(X_i)_{cj} - (\mu_X)_{cj}][(X_i)_{cj} - (\mu_X)_{cj}]^T \\ &= \sum_{i=1}^{s} A_i A_i^T \end{aligned} \qquad \textbf{(10)}$$

Now, we show that 2DPCA also finds the projection that maximizes the sum of all squared pair-wise distances between the projected data .

Theorem 2 . *2DPCA computes the m-dimensional project that maximizes*

$$\sum_{i<j} d(Y_i, Y_j)^2 \qquad \textbf{(11)}$$

Proof. By Lemma 1, we get

$$\begin{aligned} tr(W^T G_T W) &= \frac{1}{N} tr(\sum_{i=1}^{s} W^T A_i L^u A_i^T W) \\ &= \frac{1}{N} tr(\sum_{i=1}^{s} B_i L^u B_i^T) \\ &= \frac{1}{N}\sum_{l=1}^{s}\sum_{i<j} d((Y_i)_{cl}, (Y_j)_{cl})^2 \\ &= \frac{1}{N}\sum_{i<j} d(Y_i, Y_j)^2 \end{aligned} \qquad \textbf{(12)}$$

Proof is done.

Formulating 2DPCA as in (11) implies a straightforward generalization—simply replace the unit Laplacian with a general one in the target function. In the notation of Theorem 2, this means that the m-dimensional projection will maximize a weighted sum of squared distances, instead of an unweighted sum. Hence, it would be natural to call such a projection method by the name 2D Weighted PCA (2DWPCA).

Let us formalize this idea. Let be $\{wt_{ij}\}_{i,j=1}^{N}$ symmetric nonnegative pair-wise weights, with measuring how important it is for us to place the data elements i and j further apart in the low dimensional space. By convention, $wt_{ij} = 0$ for $i = j$. For this reason, these weights will be called dissimilarities in the context of weighted PCA. Normally, they are either supplied from an external source, or calculated from the data coordinates, in order to reflect any desired relationships between the data elements.

Let define NxN Laplacian $L^w_{ij} = \begin{cases} \sum_{i \neq j} wt_{ij} & i = j \\ -wt_{ij} & i \neq j \end{cases}$

and $wt_{ij} = \begin{cases} 0 & x_i, x_j \in same \quad class \\ 1/d(x_i, x_j) & other \end{cases}$

Proposition 2 . *The m-dimensional project that maximizes*

$$\sum_{i<j} w_{ij} d(Y_i, Y_j)^2 \qquad \textbf{(13)}$$

is obtained by taking the direction vectors to be the m highest eigenvectors of the matrix $\sum_{i=1}^{s} A_i L^w A_i^T$.

Proof. By Lemma 1 and 2, we get

$$\begin{aligned} & tr(W^T(\sum_{i=1}^{s} A_i L^w A_i^T)W) = tr(\sum_{i=1}^{s} W^T A_i L^w A_i^T W) \\ &= \frac{1}{N} tr(\sum_{i=1}^{s} B_i L^w B_i^T) = \frac{1}{N}\sum_{l=1}^{s}\sum_{i<j} w_{ij} d((Y_i)_{cl}, (Y_j)_{cl})^2 \\ &= \sum_{i<j} w_{ij} d(Y_i, Y_j)^2 \end{aligned} \qquad \textbf{(14)}$$

Proof is done. The 2DWPCA seeks for the m-dimensional projection that maximizes $\sum_{i<j} wt_{ij} d(Y_i, Y_j)^2$. And this is obtained by taking the m highest eigenvectors of the matrix $\sum_{i=1}^{s} A_i L^w A_i^T$.

IV. EXPERIMENTAL RESULTS

This section evaluates the performance of our propoped algorithm 2DWPCA compared with that of the 2DPCA algorithm based on using ORL and Yale face databases. 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). 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,

TABLE I
The recognition rates with 2DPCA and 2DWPCA on ORL database

d	2		4		6		8		10	
k	2DPCA	2DWPCA	2DPCA	2DWPCA	2DPCA	2DWPCA	2DPCA	2DWPCA	2DPCA	2DWPCA
2	41.56	**43.95**	59.33	**63.37**	67.48	**70.18**	71.93	**74.44**	77.11	**79.14**
3	43.5	**46.17**	75.17	**78.89**	79.33	**81.62**	82.67	**85.47**	87.67	**91.49**
4	44.1	**54.2**	72.67	**74.11**	84.1	**88.13**	89.81	**91.72**	91.71	**95.06**
5	58.22	**60.3**	73.78	**76.01**	84.89	**85.55**	88.22	**89.92**	89.33	**92.77**

TABLE II
The recognition rates with 2DPCA and 2DWPCA on Yale database

d	2		4		6		8		10	
k	2DPCA	2DWPCA	2DPCA	2DWPCA	2DPCA	2DWPCA	2DPCA	2DWPCA	2DPCA	2DWPCA
2	44.69	**49.24**	66.56	**67.11**	74.69	**76.22**	83.13	**86.35**	83.49	**87.05**
3	45.36	**49.84**	71.79	**73.49**	75	**77.75**	83.21	**87.09**	85.36	**87.72**
4	43.75	**46.62**	68.75	**72.86**	83.33	**87.35**	88.75	**90.76**	91.25	**94.03**
5	42	**46.33**	73	**77.57**	84.5	**89.57**	90.5	**93.97**	94	**96.39**

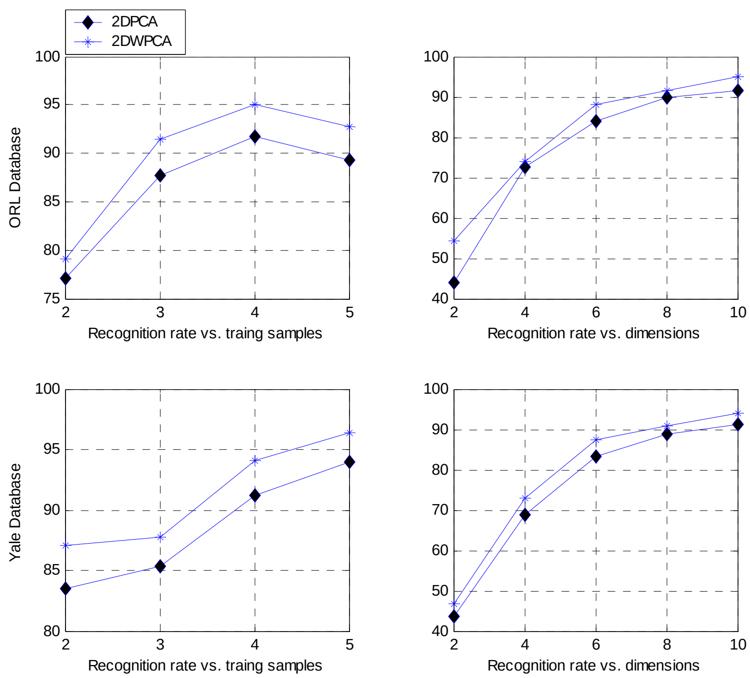

Fig. 1. The recognition rate (%) graphs which compare 2DPCA & 2DWPCA based on ORL and Yale databases

with/without glasses, normal, right-light, sad, sleepy, surprised, and wink.

In our experiments, we tested the recognition rates with different number of training samples. $k(k=2,3,4,5)$ images of each subject are randomly selected from the database for training and the remaining images of each subject for testing. For each value of k , 30 runs are performed with different random partition between training set and testing set. And for each k training sample experiment, we tested the recognition rates with different number of dimensions , d , which are from 2 to 10.

Table I&II shows the average recognition rates (%) with ORL and Yale database. In Fig. 1, we plot the graphs to make us see the recognition results of those methods intuitively. Two upper graphs are performed on ORL database, while the two lower ones are evaluated with Yale database. In recognition rate vs. training samples test, we choose the dimension d=10, and in recognition rate vs. dimension test, we choose the training sample k=4. We can see that our method achieves the better recognition rate compared to the 2DPCA.

V. Conclusions

A new 2DPCA-based method for face recognition has been proposed in this paper. The proposed 2DPCA-based method can outperform the 2DPCA method. Both PCA and

2DPCA may fail to emphasize the discrimination between the clusters, no matter how easy the task is, as they are unsupervised techniques. The directions that maximize the scatter of the data might not be as adequate to discriminate between clusters. So we proposed new 2DPCA-based schemes which can straightforwardly take into consideration data labeling, and makes the performance of recognition system better. The effectiveness of the proposed approach can be seen through our experiments based on ORL and Yale face databases. This approach can improve the performance of 2DPCA approach whose complexity is much less than PCA, LDA, or ICA approaches.

REFERENCES

[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] A. Pentland, *"Looking at People: Sensing for Ubiquitous and Wearable Computing,"* IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 107-119, Jan. 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,"* Proc. Int'l Neural Network Conf., pp. 322-325, 1990.

[7] D. Valentin, H. Abdi, A.J. O'Toole, and G.W. Cottrell, *"Connectionist Models of Face Processing: a Survey,"* Pattern Recognition, vol. 27, no. 9, pp. 1209-1230, 1994.

[8] P.S. Penev and L. Sirovich, *"The Global Dimensionality of Face Space,"* Proc. Fourth IEEE Int'l Conf. Automatic Face and Gesture Recognition, pp. 264- 270, 2000.

[9] L. Zhao and Y. Yang, *"Theoretical Analysis of Illumination in PCA-Based Vision Systems,"* Pattern Recognition, vol. 32, no. 4, pp. 547-564, 1999.

[10] L. Wiskott, J.M. Fellous, N. Kru¨ ger, and C. von der Malsburg, *"Face Recognition by Elastic Bunch Graph Matching,"* IEEE Trans. Pattern Analysis and Machine Intelligence, 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 Trans. Neural Networks, vol. 13, no. 6, pp. 1450-1464, 2002.

[12] B.A. Draper, K. Baek, M.S. Bartlett, J.R. Beveridge, *"Recognizing Faces with PCA and ICA,"* Computer Vision and Image Understanding: special issue on face recognition, in press.

[13] P.C. Yuen and J.H. Lai, *"Face Representation Using Independent Component Analysis,"* Pattern Recognition, vol. 35, no. 6, pp. 1247-1257, 2002.

[14] M.H. Yang, *"Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods,"* Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition (RGR'02), pp. 215-220, May 2002.

[15] Koren, Y.; Carmel, L. **"***Robust linear dimensionality reduction***"** Visualization and Computer Graphics, IEEE Transactions on , Volume: 10 , Issue: 4 , July-Aug. 2004 Pages:459 – 470

[16] Jian Yang; Zhang, D.; Frangi, A.F.; Jing-yu Yang "*Two-dimensional PCA: a new approach to appearance-based face representation and recognition*" Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume: 26 , Issue: 1 , Jan 2004 Pages:131 – 137

[17] "*The ORL face database*" http://www.uk.research.att.com/facedatabase.html. "*Yale face database*" http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

[18] K. Fukunaga, "*Introduction to statistical pattern recognition*", Academic Press, new York, 2 edition, 1990.