An Improvement on PCA Algorithm for Face Recognition

Vo Dinh Minh Nhat and Sungyoung Lee

Kyung Hee University, South of Korea
{vdmnhat,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. Despite these efforts, there persist in the traditional PCA some weaknesses. In this paper, we propose a new PCA-based method that can overcome one drawback existed in the traditional PCA method. In face recognition where the training data are labeled, a projection is often required to emphasize the discrimination between the clusters. PCA 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 PCA-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 traditional PCA method.

1 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]. 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. 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.

J. Wang, X. Liao, and Z. Yi (Eds.): ISNN 2005, LNCS 3496, pp. 1016–1021, 2005.

In face recognition where the data are labeled, a projection is often required to emphasize the discrimination between the clusters. PCA 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 PCA scheme 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 traditional PCA 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 the Yale face image databases to demonstrate the effectiveness of our method. Finally, conclusions are presented in Section 5.

2 Principle Component Analysis

Let us consider a set of *N* sample images $\{x_1, x_2, ..., x_N\}$ taking values in an *n*-dimensional image space, and the matrix $A = [\overline{x_1}\,\overline{x_2}...\overline{x_N}] \in \mathbb{R}^{nxN}$ with $\overline{x_i} = x_i - \mu$ and $\mu \in \mathbb{R}^n$ is the mean image of all samples. 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 $y_k \in \mathbb{R}^m$ are defined by the following linear transformation:

$$y_k = W^T \overline{x_k} \quad \text{and} \quad Y = W^T A \tag{1}$$

where $k = 1, 2, ..., N$ and $W \in \mathbb{R}^{nxm}$ is a matrix with orthonormal columns.

If the total scatter matrix is defined as

$$S_T = AA^T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \tag{2}$$

In PCA, the projection W_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\max_W \left| W^T S_T W \right| = [w_1 w_2 ... w_m] \tag{3}$$

where $\{w_i \,|\, i = 1, 2, ..., m\}$ is the set of *n*-dimensional eigenvectors of S_T corresponding to the *m* largest eigenvalues.

3 Our Proposed PCA

In the following part, we show that PCA finds the projection that maximizes the sum of all squared pairwise distances between the projected data elements and we also propose our approach. Firstly we will take a look at some necessary background.

The *Laplacian* is a key entity for describing pairwise relationships between data elements. This is an symmetric positive-semidefinite matrix, characterized by having zero row and column sums. Let L be an NxN Laplacian and $z = [z_1 z_2 ... z_N]^T \in \mathbb{R}^N$ then we have

$$\begin{aligned} z^T L z &= \sum_i L_{ii} z_i^2 + 2\sum_{i<j} L_{ij} z_i z_j = \\ &= \sum_{i<j} -L_{ij}(z_i^2 + z_j^2) + 2\sum_{i<j} L_{ij} z_i z_j = \sum_{i<j} -L_{ij}(z_i - z_j)^2 \end{aligned} \tag{4}$$

Let $r_1, r_2, ..., r_m \in \mathbb{R}^N$ be m columns of the matrix Y^T, applying (4) we have

$$\sum_{k=1}^{m} r_k^T L r_k = \sum_{i<j} -L_{ij}\left(\sum_{k=1}^{m}((r_k)_i - (r_k)_j)^2\right) = \sum_{i<j} -L_{ij} d(y_i, y_j)^2 \tag{5}$$

with $d(y_i, y_j)$ is the Euclidean distance. Now we turn into proving the following theorem, and develop it to our approach.

Theorem 1. *PCA computes the m-dimensional project that maximizes*

$$\sum_{i<j} d(y_i, y_j)^2 \tag{6}$$

Before proving this Theorem, we define a NxN unit Laplacian, denoted by L^u, as $L^u = N\delta_{ij} - 1$. We have

$$AL^u A^T = A(NI_N - U)A^T = NS_T - AUA^T = NS_T \tag{7}$$

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. By (5), we get

$$\sum_{i<j} d(y_i, y_j)^2 = \sum_{i=1}^{m} y_i^T L^u y_i = \sum_{i=1}^{m} w_i^T AL^u A^T w_i = N\sum_{i=1}^{m} w_i^T S_T w_i \tag{8}$$

Formulating PCA as in (6) implies a straightforward generalization - simply replace the unit Laplacian with a general one in the target function. In the notation of Theorem 1, 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 weighted PCA. Let us formalize this idea. Let be $\{wt_{ij}\}_{i,j=1}^{N}$ symmetric nonnegative pairwise weights, with measuring how important it is for us to place the data elements i and j further apart in the low dimensional space. Let define NxN Laplacian $L_{ij}^w = \begin{cases} \sum_{i \neq j} wt_{ij} & i = j \\ -wt_{ij} & i \neq j \end{cases}$ and $wt_{ij} = \begin{cases} 0 & x_i, x_j \in same \quad class \\ 1/d(x_i, x_j) & other \end{cases}$. Generalizing (7), we have weighted PCA and it 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 $AL^w A^T$. The proof of

this is the same as that of Theorem 1, just replace L^u by L^w. Now, we still have one thing need solving. It is how to get the eigenvectors of $AL^w A^T \in \mathbb{R}^{nxn}$, because this is a very big matrix. And the other one is how to define wt_{ij}. Let D be the N eigenvalues diagonal matrix of $A^T AL^w \in \mathbb{R}^{NxN}$ and V be the matrix whose columns are the corresponding eigenvectors, we have

$$A^T AL^w V = VD \Leftrightarrow AL^w A^T (AL^w V) = (AL^w V)D \tag{9}$$

From (9), we see that $AL^w V$ is the matrix whose columns are the first N eigenvectors of $AL^w A^T$ and D is the diagonal matrix of eigenvalues.

4 Experimental Results

This section evaluates the performance of our propoped algorithm compared with that of the original PCA algorithm and proposed algorithm (named WPCA) based on using ORL and Yale face image database. In our experiments, firstly 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 samples experiment, we tested the recognition rates with different number of dimensions, d, which are from 2 to 10. *Table 1& 2* shows the average recognition rates (%) with ORL database and Yale database respectively. In Fig. 1, we can see that our method achieves the better recognition rate compared to the traditional PCA.

Table 1. The recognition rates on ORL database.

d	2		4		6		8		10	
k	PCA	WPCA	PCA	WPCA	PCA	WPCA	PCA	WPCA	PCA	WPCA
2	39.69	**44.24**	61.56	**62.11**	69.69	**71.22**	78.13	**81.35**	78.49	**82.05**
3	40.36	**44.84**	66.79	**68.49**	70.00	**72.75**	78.21	**82.09**	80.36	**82.72**
4	38.75	**41.62**	63.75	**67.86**	78.33	**82.35**	83.75	**85.76**	86.25	**89.03**
5	37.00	**41.33**	68.00	**72.57**	79.50	**84.57**	85.50	**88.97**	89.00	**91.39**

Table 2. The recognition rates on Yale database.

d	2		4		6		8		10	
k	PCA	WPCA	PCA	WPCA	PCA	WPCA	PCA	WPCA	PCA	WPCA
2	40.56	**42.95**	58.33	**62.37**	66.48	**69.18**	70.93	**73.44**	76.11	**78.14**
3	42.50	**45.17**	74.17	**77.89**	78.33	**80.62**	81.67	**84.47**	86.67	**90.49**
4	43.10	**53.20**	71.67	**73.11**	83.10	**87.13**	88.81	**90.72**	90.71	**94.06**
5	57.22	**59.30**	72.78	**75.01**	83.89	**84.55**	87.22	**88.92**	88.33	**91.77**

5 Conclusions

A new PCA-based method for face recognition has been proposed in this paper. The proposed PCA-based method can overcome one drawback existed in the traditional PCA method. PCA may fail to emphasize the discrimination between the clusters, no

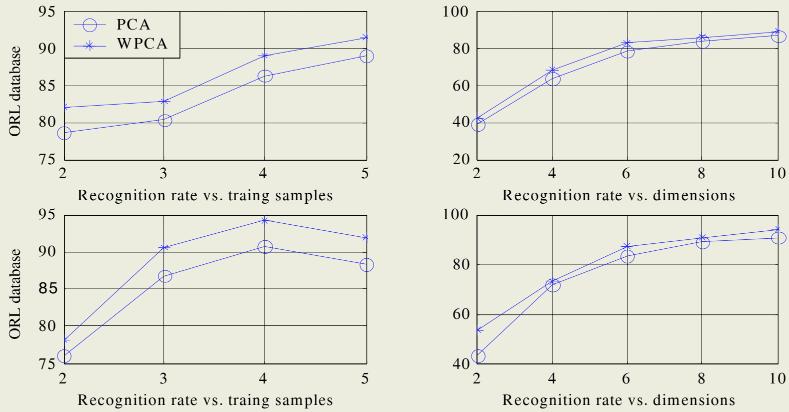

Fig. 1. The recognition rate (%) graphs on two databases.

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 PCA-based scheme 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. Perhaps, this approach is not a novel technique in face recognition, however it can improve the performance of traditional PCA approach whose complexity is less than LDA or ICA approaches.

Reference

1. Sirovich, L., Kirb, M.: Low-Dimensional Procedure for Characterization of Human Faces. J. Optical Soc. Am, **4** (1987) 519-524
2. Kirby, M., Sirovich, L.: Application of the KL Procedure for the Characterization of Human Faces. IEEE Trans. Pattern Analysis and Machine Intelligence, **12** (1990) 103-108
3. Turk, M., Pentland, A.: Eigenfaces for Recognition. J. Cognitive Neuroscience, **3** (1991) 71-86
4. Pentland, A.: Looking at People: Sensing for Ubiquitous and Wearable Computing. IEEE Trans. Pattern Analysis and Machine Intelligence, **22** (2000) 107-119
5. Grudin, M.A.: On Internal Representations in Face Recognition Systems. Pattern Recognition, **33** (2000) 1161-1177
6. Cottrell, G.W., Fleming, M.K.: Face Recognition Using Unsupervised Feature Extraction. Proc. Int'l Neural Network Conf., (1990) 322-325
7. Valentin, D., Abdi, H., O'Toole, A.J., Cottrell, G.W.: Connectionist Models of Face Processing: A Survey. Pattern Recognition, **27** (1994) 1209-1230
8. Penev, P.S., Sirovich, L.: The Global Dimensionality of Face Space. Proc. Fourth IEEE Int'l Conf. Automatic Face and Gesture Recognition, (2000) 264- 270
9. Zhao L., Yang, Y.: Theoretical Analysis of Illumination in PCA-Based Vision Systems. Pattern Recognition, **32** (1999) 547-564

10. Wiskott, L., Fellous, J.M., Kruger, N., Von Der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. IEEE Trans. Pattern Analysis and Machine Intelligence, **19** (1997) 775-779
11. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face Recognition by Independent Component Analysis. IEEE Trans. Neural Networks, **13** (2002) 1450-1464
12. Draper, B.A., Baek, K., Bartlett, M.S., Beveridge, J.R.: Recognizing Faces with PCA and ICA. Computer Vision and Image Understanding: Special Issue on FaceRecognition
13. Yuen, P.C., Lai, J.H.: Face Representation Using Independent Component Analysis. Pattern Recognition, **35** (2002) 1247-1257
14. Yang M.H.: Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods. Proc Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition (RGR'02), (2002) 215-220
15. Koren Y., Carmel L.: Robust linear dimensionality reduction. Visualization and Computer Graphics, IEEE Transactions, **10** (2004) 459-470