

Fuzzy Kernel Two-Dimensional Principal Component Analysis for Face Recognition

J.X. Zeng, P. Chen, J.Q. Tian, X. Fu

Computer Vision Institute, Nanchang Hangkong University
Hangkong, China

Abstract — The traditional kernel two-dimensional principal component analysis (K2DPCA) method did not take full advantage of the class information for face images and there are both “outer class” problem and “hard classifier” problem on face recognition. Therefore, a new face recognition method based on fuzzy kernel two-dimensional principal component analysis (FK2DPCA) is presented. Firstly, it introduces fuzzy concept into K2DPCA. Secondly, the class separability of criterion will be extended to high dimensional feature space by the use of kernel method. Furthermore, we select the eigenvectors that between-class scatter is greater than within-class scatter after projection as optimal projection axis. Finally, it uses the nearest neighbor classifier for face recognition. The experiment results on ORL and YALE face databases show that the FK2DPCA is better than other traditional methods.

Keywords—face recognition; kernel two-dimensional principal component analysis (K2DPCA); fuzzy; class separability criterion

I. INTRODUCTION

Face recognition has become an important research direction of biometrics, face feature description is one of the key steps[1], mainly by extracting the effective face identification and the design of complex classifier[2]. Principal component analysis (PCA)[3] and two dimensional principal component analysis of (2DPCA)[4] are both the linear feature extraction method of minimum square error sense. However face space is a nonlinear manifold structure, they cannot effectively extract nonlinear face structure feature for classification. Kernel techniques can effectively to the traditional subspace methods for nonlinear improvement. Hui Kong et proposed a two-dimensional kernel principal component analysis (K2DPCA) method[5], which uses the kernel learning method not only to extract nonlinear features of human face but also to map the nonlinear inseparable face image into the high-dimensional feature space H , establishing the optimal hyper plane in H to realize the linear separable[6]. But PCA, 2DPCA and K2DPCA methods all do not make full use of class information of training samples[7]. So Li Yong zhi et proposed a combination of class information kernel principal component analysis method[8], where the training samples of known classification information are put into the feature extraction of face. But there are still two problems: one hand when the sample is far away from the average of itself by virtue of illumination, expression and accessories etc, the edge of class problem[9] will appear, in which case the sample classification is not scientific any more and there is a hard classification problem[10]; on the other hand, the method that only choose the eigenvectors corresponding to the larger eigenvalues as the optimal

projection axis[11], discards some face information easy to identify. So the selection of the optimal projection axis is not accurate.

This paper proposed a face recognition method based on fuzzy two-dimensional kernel principal component analysis (FK2DPCA). First we introduce the fuzzy concepts in K2DPCA, using fuzzy K nearest neighbor algorithm to each category of membership calculation of sample the class center in H and definite fuzzy scatter matrix H based on the membership information for the categories and distribution information of samples being fully integrated into the facial feature extraction. Some improvements aiming are made for the edge problem in face recognition produced by face illumination, facial expressions and so on. Secondly, the use of nuclear learning class separability criterion will be extended to high-dimensional feature space H . Then, we select the eigenvectors that between-class scatter is greater than within-class scatter after projection as optimal projection axis, in order to obtain the optimal discriminant features of face. At the same time the nearest neighbor classifier is used for classification and recognition.

II. TWO-DIMENSIONAL FUZZY KERNEL PRINCIPAL COMPONENT ANALYSIS METHOD

There are two fundamental problems on K2DPCA method. One is to select only the relatively larger eigenvectors as the optimal projection axis, disregarding the relatively small nonzero eigenvalues and the corresponding eigenvectors which contains face information beneficial to recognition[11]. On the other hand it does not fully utilize the face training samples of known category information.

A. Two-Dimensional Fuzzy Basic IDEA of Kernel Principal Component Analysis Method

For hard classification problems and edge issues of K2DPCA method existing in face recognition, Similar method is used in the literature[12]. First of all, we introduces the fuzzy concept into K2DPCA, using the corresponding membership functions to put the class information and the sample distribution information into feature extraction. Then we build the class separability criterion for samples in high dimensional space. Finally, we use the nearest neighbor classifier to classify recognition. The sample membership functions can be got by the rule of fuzzy neighbour[10]:

$$\mu_{i(jk)} = \begin{cases} 0.51+0.49 \times (n_{i(jk)} / K) & \text{when } i \in j \\ 0.49 \times (n_{i(jk)} / K) & \text{when } i \notin j \end{cases} \quad (1)$$

$\sum_{i=1}^c \mu_{i(jk)} = 1$, $0 < \sum_{i=1}^c \sum_{k=1}^{n_i} \mu_{i(jk)} < N$ and $\mu_{i(jk)}$ is the dependence of k samples of the class j for the class i , $n_{i(jk)}$ is the number of K nearest neighbor of the k samples of class j belonging to the class i .

First of all we make up the distance matrix D_1 by the distance between the training samples. Second, we set the second diagonal matrix D_1 to infinity; then by size order matrix of each line, we can get the matrix D_2 , by which K nearest neighbor points and the corresponding category information can be obtained. Finally by (1), the corresponding membership degree matrix $\mu = [\mu_{i(jk)}]$ of the training sample is calculated in the \mathbf{H} .

The sample average of the i th sample in high dimensional feature space \mathbf{H} can be got by using the fuzzy operation Fu_i .

We set N face training samples of Class c as \mathbf{A} , and $A_i^k \in \mathbf{A}$, Where A_i^k is the k sample images of class i . n_i is the i th training sample number of sample B_i , and N is the total number of training samples. $\varphi(A_i^k)$ is the nuclear sample matrix of sample A_i^k got by the nonlinear transformation function φ being mapped to high dimensional feature space \mathbf{H} (The dimension of feature space \mathbf{H} is L). $\overline{\varphi(A_i)}$ is the corresponding average of i th class sample in \mathbf{H} . $\overline{\varphi(\mathbf{A})}$ is the mean corresponding to of all the training samples in \mathbf{H} , and P_i is the priori probability of i th class sample, which is defined as follows:

Definition 1: the fuzzy within-class scatter matrix for the sample in \mathbf{H} is:

$$FS_w = \sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^{n_j} \mu_{i(jk)}^m (\varphi(A_j^k) - Fu_i)(\varphi(A_j^k) - Fu_i)^T \quad (2)$$

Definition 2: the fuzzy between-class scatter matrix for the sample in \mathbf{H} is:

$$FS_b = \sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^{n_j} \mu_{i(jk)}^m (Fu_i - Fu)(Fu_i - Fu)^T \quad (3)$$

Fu is the overall sample average. m is the fuzzy index.

Definition 3: the fuzzy total scatter matrix for the sample in \mathbf{H} is:

$$FS_t = \sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^{n_j} \mu_{i(jk)}^m (\varphi(A_j^k) - Fu)(\varphi(A_j^k) - Fu)^T \quad (4)$$

Corollary 1: Samples in \mathbf{H} fuzzy overall divergence in the matrix FS_t is samples in \mathbf{H} fuzzy divergence between class

matrix and fuzzy divergence within class FS_b matrix FS_w combined.

B. Class Separability Criterion in High Dimensional Feature Space

From the perspective of selecting the optimal projection axis, the kernel-based learning method is introduced into the class separability criterion[13] forming the high dimensional feature space \mathbf{H} class separability criterion, to improve the accuracy of selecting the optimal projection axis.

Different face samples in the original image in the feature space is not necessarily linear separable, but is linearly separable in high dimensional feature space. In order to achieve the non-linear separate original face image in the feature space and the linearly separable in the high dimensional feature space \mathbf{H} , K2PDCA face recognition methods, first of all, use the nuclear method to map face image into \mathbf{H} by nonlinear transformation. Then we can get the corresponding optimal projection axis, which is used to establish the most optimal projection space. Finally when face image feature can be represented by a point on the \mathbf{H} , we compute the distance between two points in the two-dimensional space method to calculate the difference between two samples, and the size of the difference can be measured by the size of the available distance as class separability criterion in high dimensional feature space.

Therefore, putting the class separability criterion into high-dimensional feature space can get available class separability criterion in high dimensional feature space.

Definition 4: the between-class scatter matrix for the sample in \mathbf{H} is:

$$S_b^w = \sum_{i=1}^c P_i (\overline{\psi(A_i)} - \overline{\psi(\mathbf{A})})(\overline{\psi(A_i)} - \overline{\psi(\mathbf{A})})^T \quad (5)$$

Definition 5: the within-class scatter matrix for the sample in \mathbf{H} is:

$$S_w^w = \sum_{i=1}^c P_i \frac{1}{n_i} \sum_{j=1}^{n_i} (\psi(A_i^j) - \overline{\psi(A_i)})(\psi(A_i^j) - \overline{\psi(A_i)})^T \quad (6)$$

Definition 6: the overall scatter matrix for the sample in \mathbf{H} is

$$S_t^w = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^{n_i} (\psi(A_i^j) - \overline{\psi(\mathbf{A})})(\psi(A_i^j) - \overline{\psi(\mathbf{A})})^T = S_b^w + S_w^w \quad (7)$$

$\overline{\psi(A_i)}$ is the i th sample average in \mathbf{H} and $\overline{\psi(\mathbf{A})}$ is the average of all the training samples in \mathbf{H} . The trace of overall scatter matrix trace available is :

$$J = \overline{dist^2(\psi(\mathbf{A}))} = Tr[S_w^w + S_b^w] = Tr[S_t^w] \quad (8)$$

So the trace J of the total scatter matrix S_i^o can be treated as class separability criterion in sample H . The greater J stands for the more dispersed samples.

C. Feature Vector Selection and Face Recognition

Due to the relatively smaller non-zero eigenvalues and the corresponding eigenvectors in K2DPCA method does not participate in the optimal projection axis, which leads to lose a part of the face information for identification[11]. So this paper makes some improvements.

Firstly, according to the high dimensional feature space separability criterion in H , we can get all non-zero eigenvalue projection space $w = \{w_1, w_2, \dots, w_h\}$, where h is the number of non-zero eigenvalue in FS_i .

Secondly, we choose the one the between class scatter being larger than the within class scatter after projection as the optimal projection axis.

$$w_i^T FS_b w_i > w_i^T FS_w w_i \quad (9)$$

so the optimal discriminant features of face can be expressed as:

$$Y = w_{FK2DPCA}^T \varphi(A) \quad (10)$$

where $w_{FK2DPCA} = [w_1, w_2, \dots, w_i, \dots, w_d]$ represents the optimal projection space, and w_i is an optimal projection axis in the optimal projection space. d is the optimal projection axis number, $d \leq h$, and $\varphi(A)$ represents a collection of the corresponding training samples of H . By the formula (10), we can obtain the optimal characteristics of the A_i^k face image of the p individual identification:

$$Y_p = w_{FK2DPCA}^T \varphi(A_i^k) \quad (11)$$

By using the same method, we can also get the corresponding optimal identifying characteristics Y_{test} . Then, the nearest neighbor classifier is used for classification and recognition, i.e.

$$dist(Y_p, Y_{test}) = \min_{p=1}^N dist(Y_p, Y_{test}) \quad (12)$$

We can get that when Y_p belongs to the i class samples, test sample Y_{test} belongs to the i class of face images.

III. EXPERIMENTAL RESULTS AND ANALYSES

In order to verify the effectiveness of the algorithm for PCA, 2DPCA, K2DPCA, FK2DPCA face recognition methods are in ORL face database and YALE comparative experiments carry out on.

In order to make the experiment more comparable, 1,2,...,9 face images are selected as a training sample by order, and we make tests using the remaining samples. To reduce the calculation amount, the image dimensions ORL and YALE face image database were normalized to 28×24 and 61×80 , and the kernel function is Gaussian kernel function: $k(x, y) = \exp\{\|x - y\|^2 / \sigma\}$, $\sigma = 9 \times 10^5$, and the nearest neighbor classifier is used for face classification and recognition.

From table 1. and table 2., figure 1. and Figure 2. can see, the FK2DPCA method is more stable and efficient than PCA, 2DPCA and K2DPCA on the overall performance face recognition. This is because that the fuzzy membership information of the sample class information and the distribution of information is completely integrated into the final feature extraction in FK2DPCA method and we make effective improvement on the edge and hard classification problem existing in face recognition. Then we select the between class scatter being larger than the within class scatter as the optimal projection axis after projection, which may be smaller feature selection with face identification features of the value of the eigenvectors corresponding to participate in the optimal projection axis.

TABLE I. THE AVERAGE RECOGNITION RATE AMONG FOUR METHODS ON ORL FACE DATABASE

Method	1	2	3	4	5	6	7	8	9
PCA	0.631	0.716	0.754	0.793	0.829	0.865	0.907	0.902	0.897
2DPCA	0.711	0.820	0.852	0.865	0.900	0.940	0.943	0.946	0.934
K2DPCA	0.731	0.899	0.901	0.883	0.911	0.936	0.947	0.947	0.941
FK2DPCA	0.730	0.892	0.915	0.908	0.914	0.948	0.948	0.953	0.952

TABLE II. THE AVERAGE RECOGNITION RATE AMONG FOUR METHODS ON YALE FACE DATABASE

Method	1	2	3	4	5	6	7	8	9
PCA	0.395	0.483	0.596	0.635	0.633	0.621	0.640	0.644	0.642
2DPCA	0.504	0.598	0.725	0.731	0.748	0.731	0.755	0.750	0.756
K2DPCA	0.520	0.633	0.716	0.741	0.775	0.767	0.802	0.801	0.813
FK2DPCA	0.523	0.679	0.734	0.746	0.781	0.783	0.802	0.802	0.817

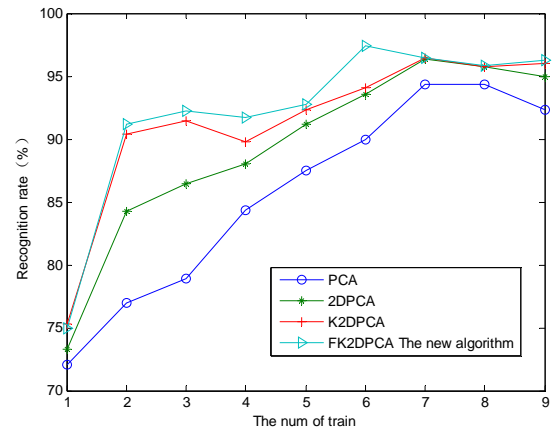


FIGURE I. THE BEST RECOGNITION RATE COMPARISON ON ORL

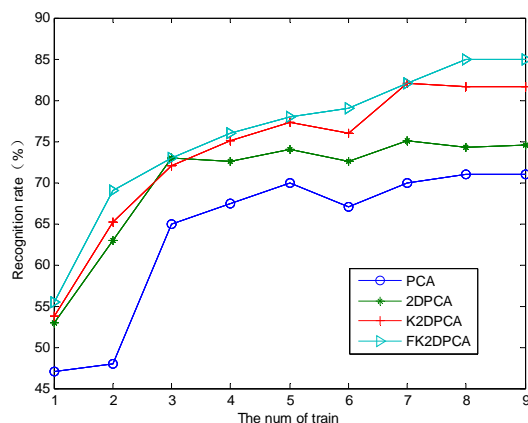


FIGURE II. THE BEST RECOGNITION RATE COMPARISON ON YALE

IV. SUMMARY

Theory analysis and experimental results show that the algorithm in this paper, compared to many of the traditional algorithm, has certain advantages on the face recognition. First of all, face recognition method of fuzzy two-dimensional kernel principal component analysis inherit the advantages from the K2DPCA. Then, the fuzzy concept is introduced into face recognition method of K2DPCA. We make effective improvements in this paper, defining the samples in the high dimensional feature space fuzzy scatter matrix and testing center by membership information, putting the distribution information of samples and the categories of information completely into the feature extraction in the last class. What is more, we take the edge of the class of problems and the presence of the hard classification for Face vulnerable illumination, expression and other factors into consideration. Finally, we establish separability criterion of face samples in the high dimensional space and select the eigenvectors that between-class scatter is greater than within-class scatter after projection as optimal projection axis, to improve the accuracy of the method and the expression ability for facial features. The experiments demonstrate the effectiveness of this algorithm.

ACKNOWLEDGMENTS

This work was supported partially by the the National Natural Science Foundation of China (Grant No. 61165011, 61263046), the Youth Science Foundation of Jiangxi province (Grant No. 20132BAB211021), the Scientific research project Department of education of Jiangxi province, China (Grant No. GJJ12427).

REFERENCES

- [1] Vun S., Dee H. M., Caplier A. Face Recognition Using the Poem Descriptor. *Pattern Recognit.*, 2012, 45: 2478-2488.
- [2] Zhang L., Zhou W. D., Chang P. C., Liu J., Yan Z., Wang T., Li F. Z. Kernel Sparse Representation based Classifier. *IEEE Transactions on Signal Processing*, 2012, 60(4): 1684-1695.
- [3] Turk M., Pentland A. Eigenfaces for Recognition. *Cognitive Neuroscience*, 1991, 3(1): 71-86.
- [4] Yang J., Zhang D., Frangi A. F., et. Two-dimensional PCA: a New Approach to Appearance-based Face Representation and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(1): 131-137.
- [5] Kong H., Wang L., Eam K. T., et al. Generalized 2D Principal Component Analysis for Face Image Representation and Recognition. *Neural Networks*, 2005, 18(5-6): 585-594.
- [6] Di Y., Hamsici O. C., Martinez A. M. Kernel Optimization in Discriminant Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(3): 631-638.
- [7] Chen Songcan, Sun Tingkai. Class Information- Incorporated Principal Component Analysis. *Neurocomputing*, 2005, 69(1-3): 216-223.
- [8] Li Yongzhi, Yang Jingyu, Wu Songsong. A Class-Information-Incorporated Kernel Principal Component Analysis Method. *Pattern recognition and Artificial intelligence*, 2008, 21(3): 410-416.
- [9] Zhuang Zhemin, Zhang Aniu, Li Fenlan. Based on an Optimized LDA Algorithm for Face Recognition. *Journal of Electronics and Information Technology*, 2007, 29(9): 2047-2049.
- [10] Yang Wankou, Wang Jianguo, Ren Mingwu, et al. Fuzzy Inverse FDA and Its Application to Face Recognition. *Journal of Image and Graphics*, 2009, 14(1): 88-93.
- [11] Ju Shenggen, Zhou Jiliu, Wang Chaobin, et. Face Recognition on Feature Fusion. *Journal of Sichuan University*, 2009, 46(3): 618-621.
- [12] Gao Jun, Wang Shitong. Fuzzy Maximum Scatter Difference Discriminant Criterion Based Clustering Algorithm. *Journal of Software*, 2009, 20(11): 2939-2949.
- [13] Yin Hongtao, Fu Ping, Sha Xuejun. Face Recognition Based on DCT and LDA. *Acta Electronica Sinica*, 2009, 37(10): 2211-2214.