ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 4, No. 3, 2009, *pp*. 233-239

Fuzzy 2-Dimensional FLD for Face Recognition

Wankou Yang [1, +], Jianguo Wang [1, 2], Mingwu Ren [1], Jingyu Yang [1]

[1]partment of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, People's Republic of China

[2]Tangshan College, Network & Education Center, Tangshan, 063000, People's Republic of China

(*Received March 16, 2008, accepted August 11, 2008*)

Abstract: This paper proposes a new method of face image feature extraction, namely, the fuzzy 2DFLD (F2DFLD) based on the 2D fisher discriminant criterion and fuzzy set theory. In the proposed method, we calculate membership degree matrix by FKNN, then we incorporate the membership degree into the definition of the between-class scatter matrix and within-class scatter matrix and get the fuzzy between-class scatter matrix and fuzzy within-class scatter matrix. Experiments on the Yale, ORL and FERET face databases show that the new method can work well

Keywords: FLD, 2DFLD, Fuzzy 2DFLD, Feature extraction, Face Recognition

1. Introduction

Fisher linear discrimination (FLD) [1], sometimes known as linear discrimination analysis (LDA), has been widely used in pattern recognition for feature extraction and dimensional reduction. The objective of FLD is to find the optimal projection so that the ratio of determinants of between-class and the within-class scatter matrices of the projected samples reaches its maximum. A difficulty in using the FLD method for face recognition is the high-dimensional nature of the image rector. In order to utilize FLD for face recognition, a number of research work has been done[2-7]. The most popular method, called Fisherface, was build by D.L.Swets et al.[2] and Belhumeur et al.[3]. In their methods, PCA is first used to reduce the dimension of the original features space to $N-c$, and the classical FLD is next applied to reduce the dimension to d ($d \le c$). Obviously, in the process K-L transform, the small $c-1$ projection components have been thrown away. So some effective discriminatory information may be lost. And PCA step can't guarantee the transformed within-class scatter matrix still be not singular. Recently, 2DFLD [8,9,10], on par with 2DPCA [11] has been proposed and successfully applied for face recognition.

Regretfully, the 2DFLD assumes that the same level of typicality of each face to the corresponding class. We propose to incorporate a gradual level of assignment to class being regarded as a membership grade with anticipation that such discrimination helps improve classification results. More specifically, when operating on feature vectors resulting from PCA transformation we complete a Fuzzy K-nearest neighbor class assignment that produces the corresponding degree of class membership. By taking advantage of the technology of fuzzy sets [12], a number of studies have carried out for fuzzy pattern recognition [13-16].

The organization of this paper is as follows: In Section2, we briefly review 2DFLD and Fuzzy fisherface. In Section 3, we propose the idea and describe the new method in detail. In Section 4, experiments with face images data are presented to demonstrate the effectiveness of the new method. Conclusions are summarized in Section 5.

2. Related works

2.1. 2DFLD

Training is a process of acquiring features from available training images and storing them in a knowledge base for the purpose of recognition an unknown future scene image. Given a set of samples of each class, the 2DFLD approach extracts most informative features which could establish a high degree of

[+] Corresponding author. Tel: +86 25 84317297-403.
E-mail: yangwankou@njust.edu.cn, wankou_yang@yahoo.com.cn.

similarity between samples of the same class and a high degree of dissimilarity between samples of two classes.

Suppose there are c known pattern classes $w_1, w_2, \cdots w_c$, N training samples. $A=\{A_j{}^i\}$ ($i=1,2,\cdots l_c$, $j=1,2,\cdots c$) is a set of samples with $m \times n$ dimension. l_j is the number of training samples of class j and satisfies $\sum_{i=1}^{c} l_i = N$. Set $\overline{A_j} = (1/l_j)\sum_{i=1}^{l_j} A_j^i$, $\overline{A} = (1/N)\sum_{j=1}^{c}\sum_{i=1}^{l_j} A_j{}^i$, $j=1,2,\cdots c$. The image between-class scatter matrix G_b and the image within-class scatter matrix G_w are computed as

$$G_b = \frac{1}{N}\sum_{j=1}^{c} l_j(\overline{A_j}-\overline{A})(\overline{A_j}-\overline{A})^T \tag{1}$$

$$G_w = \frac{1}{N}\sum_{j=1}^{c}\sum_{i=1}^{l_i}(A_j{}^i-\overline{A_j})(A_j{}^i-\overline{A_j})^T \tag{2}$$

Once G_b and G_w are computed, it is recommended to find the optimal projection axis W so that the total scatter of the projected samples of the training images is maximized. The objective function of 2DFLD can be defined as:

$$J(W) = \arg\max_W \frac{W^T G_b W}{W^T G_w W} \tag{3}$$

It can be proven that the eigenvector corresponding to the maximum eigenvalue of $(G_w)^{-1}G_b$ is the optimal projection vectors which maximizes $J(W)$. Generally, as it is not enough to have only optimal projection vector, we usually go for d number of projection axes, Say $w_1, w_2, \cdots, w_d$, which are the eigenvectors corresponding to the first d largest eigenvalues of $(G_w)^{-1}G_b$. In 2DFLD, once these projection vectors are computed, each training image A_j^i is then projected onto W to obtain the feature matrix Y_j^i of size $m \times d$ of the training image A_j^i. So, during training, for each training image A_j^i, a corresponding feature matrix of size $m \times d$, is constructed and stored in the knowledge edge for matching at the time of recognition.

2.2. Fuzzy Fisherface

K.C-Kwak [14] proposed the fuzzy fisherface for face recognition via fuzzy set. Given a set of feature vectors, $X=\{x_1, x_2, \cdots, x_{N,}\}$, a fuzzy "$c$"-class partition of these vectors specifies the degree of membership of each vector to the classes. The membership matrix $\left[u_{ij}\right](i=1,2,\cdots,c,\ j=1,2,\cdots,N)$ can be gotten by FKNN, it will be discussed in section 3.1 in detail. Taking into account the membership grades, the mean vector of each class $\overline{m_i}$ is calculated as follows:

$$m_i = \frac{\sum_{j=1}^{N} u_{ij} x_j}{\sum_{j=1}^{M} u_{ij}} \tag{4}$$

The between-class fuzzy scatter matrix SF_b and within-class fuzzy scatter matrix SF_w incorporate the membership values in their calculations

$$SF_b = \sum_{i=1}^{c} N_i(\overline{m_i}-\overline{m})(\overline{m_i}-\overline{m})^T \tag{5}$$

$$SF_w = \sum_{i=1}^{c}\sum_{x_k \in w_i}(x_k-\overline{m_i})(x_k-\overline{m_i})^T \tag{6}$$

The optimal fuzzy projection W of fuzzy Fisherface follows the expression:

$$W = \arg\max_{W} \frac{W^T SF_b W}{W^T SF_w W} \tag{7}$$

Finally, K.C-Kwak gave the strategy: PCA plus Fuzzy FLD in small sample size case.

3. Fuzzy 2DFLD

2DFLD is considered to solve binary classification problems. While overlapping samples existed, performance of 2DFLD may degenerate. How can we represent the distribution of these samples and improve classification performance through extracting discriminative information from these samples? Obviously, fuzzy set theory is a good choice. In this paper, we proposed a new Fuzzy 2DFLD (F2DFLD) algorithm, which makes fully of the distribution of samples. Samples distribution information is represented by fuzzy membership degree corresponding to every class.

3.1. Fuzzy K-Nearest Neighbor (FKNN)

In our method, fuzzy membership degree and each class center are gained through FKNN [13] algorithm. With FKNN algorithm, the computations of the membership degree can be realized through a sequence of steps:

Step1: Compute the Euclidean distance matrix between pairs of feature vectors in training set.

Step2: Set diagonal elements of this Euclidean distance matrix to infinity.

Step3: Sort the distance matrix(treat each of its columns separately) in an ascending order. Collect the corresponding class labels of the patterns located in the closest neighborhood of the pattern under consideration (as we are concerned with 'k' neighbors, this returns a list of 'k' integers).

Step4: Compute the membership degree to class 'i' for jth pattern using the expression proposed in the literature[13]

$$u_{ij} = \begin{cases} 0.51 + 0.49 \times (n_{ij}/k) & \text{if } i = \text{ the same as the } j\text{th label of the pattern} \\ 0.49 \times (n_{ij}/k) & \text{if } i \neq \text{ the same as the } j\text{th label of the pattern} \end{cases} \tag{8}$$

In the above expression n_{ij} stands for the number of the neighbors of the jth data (pattern) that belong to the ith class.

Therefore, the fuzzy membership matrix U can be achieved with the result of FKNN :

$$U = \left[u_{ij}\right],\ i = 1, 2 \cdots c,\ j = 1, 2, \cdots N. \tag{9}$$

3.2. The idea of Fuzzy 2DFLD

The key step of Fuzzy 2DFLD is how to incorporate the contribution of each training sample into the redefine of scatter matrices. With the conception of fuzzy set theory, every sample can be classified into multi classes under fuzzy membership degree, which is different to binary classification problem. In the redefinition of the fuzzy within-class scatter matrix, samples that are more close to class center have more contribution to classification. In the redefinition of the between-class scatter matrix, in contrast to the redefinition of the within-class scatter matrix, class which is far from total center will have more contribution to classification. Then, the membership degree of each sample (contribution to each class) should be considered and the mean matrix of each class, the corresponding fuzzy within-class scatter matrix and fuzzy between-class scatter matrix can be redefined as follow:

$$\overline{A_i} = \frac{\sum_{j=1}^{N} u_{ij}^{p} A_j}{\sum_{j=1}^{M} u_{ij}^{p}} \tag{10}$$

$$FG_w = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^{p} (A_j - \overline{A_i})(A_j - \overline{A_i})^T \tag{11}$$

$$FG_b = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^{p} (\overline{A_i} - \overline{A})(\overline{A_i} - \overline{A})^T \tag{12}$$

Where p is a constant which controls the influence of fuzzy membership degree, $\overline{A}$ is the mean of all samples. Similarly, the fuzzy total scatter matrix can be achieved as follows:

$$FG_t = FG_w + FG_b \tag{13}$$

So, all scatter matrices with fuzzy set theory is redefined and the contribution of each sample is incorporated. From Eq.(4-6) and Eq.(10-13), we can find that our proposed method can more make fully of the distribution of samples

Therefore, all scatter matrices with fuzzy set theory is redefined and the contribution of each sample is incorporated. It is easy to show that FG_w and FG_b are $m \times m$ matrices and often unsingular. The objective function of Fuzzy 2DFLD can be redefined as:

$$J(W) = \arg\max_W \frac{W^T FG_b W}{W^T FG_w W} \tag{14}$$

This criterion is a Rayleigh quotient in form. It is easy to find its optimal solutions by solving a generalized eigen-equation.

3.3. The algorithm of Fuzzy 2DFLD

Based on the above descriptions, Fuzzy 2DFLD (F2DFLD) algorithm can be described as follows:

Step1.(FKNN): Class center matrix $\overline{A_i}$ and fuzzy membership degree matrix U of training samples can be achieved with FKNN algorithm in original image space.

Step2.(Fuzzy 2DLDA): According to $\overline{A_i}$ and U work out fuzzy within-class scatter matrix FG_w and fuzzy between-class scatter matrix FG_b . The optimization problem (10) can be solved by $(FG_w)^{-1} FG_b \cdot u = \lambda \cdot u$ with eigenvalues $\lambda_1 > \cdots > \lambda_q > 0$, and corresponding normalized eigenvectors $u_1, u_2, \cdots u_q$. We get the optimal projection matrix.

Step3.(recognition): Project all samples into the obtained optimal discriminant matrix and classify.

4. Experiments

Three face image databases, namely, the Yale database, the ORL databse and FERET database, are used to compare the proposed Fuzzy 2DFLD approach with the following algorithms: PCA(eigenface) [3], 2DPCA [10], LDA(Fisherface) [3], 2DFLD [8], Fuzzy Fisherface [14].

4.1. Experiments Using Yale Database

The Yale face database contain 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions. In our experiments, each image was manually cropped and resized to 50×40. Figure 1 shows sample images of one person.

Figure 1. Eleven images of one person in Yale

The first experiment was performed using the first six images (i.e., center-light, with glasses, happy, left-light, without glasses, and normal) per class for training, and the remaining five images (i.e., right-light, sad, sleepy, surprised, and winking) for testing. For feature extraction, we used, respectively, PCA, 2DPCA, LDA, 2DLDA, Fuzzy Fisherface, and the proposed Fuzzy 2DLDA. Note that LDA and Fuzzy Fisherface involve a PCA phase. In this phase, we keep nearly 95 percent image energy and select the number of principal components, m , as 37. In the second phase of LDA and Fisherface, the number of discriminant vectors corresponding to the $c-1$ largest generalized eigenvalues is 14. The FKNN parameter K is set as $K = l - 1 = 5$, where l denotes the number of training samples per class. The justification for this choice is that each sample should have $l-1$ samples of the same class provided that within-class samples are well clustered. Finally, the nearest neighbor (NN) classifier with cosine distance is employed for classification. The maximal recognition rate of each method and the corresponding dimensions are given in Table1.

Table 1. Maximal Recognition Comparision On Yale

method	PCA	2DPCA	LDA	2DLDA	Fuzzy Fisherface	**proposed**
Result	0.8533	0.9200	0.9333	0.9333	0.8533	**0.9600**
Dimension	37	50×11	14	50×3	14	50×3

From Table 1, we can see that Fuzzy 2DLDA outperforms other methods. Why can Fuzzy 2DLDA outperform other methods? In our opinion, the overlapping sample's distribution information is incorporated in the redefinition of corresponding scatter matrices by fuzzy set theory, which is important for classification.

In the second experiment, 10-fold cross-validation tests are performed to reevaluate the performance of PCA, 2DPCA, LDA, 2DLDA, Fuzzy Fisherface, and Fuzzy 2DFLD. In each test, six images of each subject are randomly chosen for training, while the remaining five images are used for testing. In the second experiment, we select the same dimension and parameters as in the first experiment. Table 2 shows the maximal average recognition rates across 10 runs of each method under nearest neighbor classifier with cosine distance metrics and their corresponding standard deviations (std) and dimension. From Table 2, it can be seen that Fuzzy 2DFLD outperforms other methods.

In all the experiments, PCA is used to extract 66 principal component features and 2DPCA to extract 56×11 principal component feature vectors. LDA is used to extract 14 LDA features, 2DFLD and F2DFLD are used to extract 56×3 feature vectors.

Table 2. Recognition Comparision On Yale

method	PCA	2DPCA	LDA	2DLDA	Fuzzy Fisherface	**proposed**
Ex	0.9027	0.9240	0.9520	0.9533	0.8533	**0.9633**
Std	0.0576	0.0321	0.0566	0.0144	0.0873	**0.0261**

4.2. Experiments Using ORL Database

Figure 2. Ten images of one person in ORL

The ORL (http://www.cam-orl.co.uk) database contains 40 persons, each having 10 different images. The images of the same person are taken at different times, under slightly varying lighting conditions and with various facial expressions. Some people are captured with or without glasses. The heads in images are slightly titled or rotated. The images in the database are manually cropped and rescaled to 56×46. Figure 2 shows ten images of one person in ORL.

In the experiments, we split the whole database into two parts evenly. One part is used for training and the other part is for testing. In order to make full use of the available data and to evaluate the generalization power of algorithms more accurately, we adopt a cross-validation strategy and run the system 10 times. In each time, S face images from each person are randomly selected as training samples, and the rest is for testing. The classical PCA, 2DPCA, LDA , 2DFLD, Fuzzy Fisherface and the proposed Fuzzy 2DFLD (F2DFLD) are respectively used for feature extraction. In the PCA stage of LDA and Fuzzy Fisherface, we keep nearly 95 percent image energy and the number of principal components, m, is set as 61 and 73. The FKNN parameter K is set as $K = S - 1$. Finally a nearest neighbor classifier with cosine distance is employed. The recognition results are shown in Table 3. From Table 3, we find that our method all outperforms other methods and our method can work well.

Table 3. Recognition Comparision On ORL

method	S	Class	Dim	Mean / std
PCA	3	40	61	0.8918 / 0.0189
2DPCA	3	40	56*4	0.9004 / 0.0261
LDA	3	40	39	0.8236/ 0.0275
2DLDA	3	40	56*3	0.9161 / 0.0193
Fuzzy Fisherface	3	40-	39	0.8232 / 0.0283
proposed	**3**	**40**	**56*3**	**0.9208 / 0.0217**
PCA	4	40	73	0.9375/ 0.00224
2DPCA	4	40	56*4	0.9354 / 0.0234
LDA	4	40	39	0.8875 / 0.0268
2DLDA	4	40	56*3	0.9417 / 0.0127
Fuzzy Fisherface	4	40	39	0.8867 / 0.0227
proposed	**4**	**40**	**56*3**	**0.9504 / 0.0194**

4.3. Experiments Using FERET Database

The FERET face database is a result of the FERET program, which was sponsored by the US Department of Defense through the DARPA Program[18,19]. It has become a standard database for testing and evaluating state-of-the-art face recognition algorithms. The proposed method was tested on a subset of the FERET database. This subset includes 1,400 images of 200 individuals (each individual has seven images). This subset involves variations in facial expression, illumination, and pose. In our experiment, the facial portion of each original images was automatically cropped based on the location of the eyes and the cropped images was resized to 40×40 pixels. Some examples images of one person are shown in Figure 3.

Figure3. Seven images of one person in FERET

In order to make full use of the available data and to evaluate the generalization power of algorithms more accurately, we adopt a cross-validation strategy and run the system 10 times. In each time, 2 face images from each person are randomly selected as training samples, and the rest is for testing. The classical PCA, LDA, 2DFLD, Fuzzy Fisherface and the proposed Fuzzy 2DFLD (F2DFLD) are respectively used for feature extraction. The FKNN parameter K is set as $K = 2$. Finally a nearest neighbor classifier with cosine distance is employed. From Figure 4, we can see that F2DFLD outperforms the other methods for every number of training samples for each class. In the experiments, in the PCA stage of LDA and Fuzzy Fisherface, we keep nearly 95 percent image energy and the number of principal components, m, is set as 114. PCA is used to extract 86 principal component features, 2DFLD and F2DFLD are used to extract 40×4 feature vectors.

Table 4. Recognition Comparision On FERET

method	PCA	2DPCA	LDA	2DLDA	Fuzzy Fisherface	**proposed**
Ex	0.3662	0.4792	0.2161	0.4388	0.1964	**0.4888**
Std	0.0688	0.0864	0.0408	0.0503	0.0408	**0.0605**
Dim	114	40*8	114	40*8	114	**40*8**

The images in the FERET face database are subject to complex nonlinear changes due to large pose, expression or illumination variations. From Table 4, we can see that our proposed method can work well in the complex circumstance.

5. Conclusions

In this paper, we propose a new method for feature extraction, fuzzy 2DFLD (F2DFLD). This method is based on FLD and image matrix. Furthermore, the overlapping sample's distribution information is fully incorporated in the redefinition of corresponding scatter matrices, which is important for classification. Experiments on the Yale, the ORL and FERET face databases show that the new method can work well.

6. Acknowledgements

This project is supported by NSFC of China (Grant Nos: 60632050, 606472060, 60473039), by Hi-Tech Research and Development Program of China (Grant No. 2006AA01Z119) and by the Graduates' Research Innovation Program of Higher Education of Jiangsu Province (Grant No.CX07B_118z).

7. Reference

[1] Fisher R A. The use of multiple measurements in taxonomic problems. *Annal of Eugenics.* 1936, **7**: 179-188.

[2] Swets D L., Weng J.. Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Trans on Pattern Analysis and Machine Intelligence.* 1996, **18**(8): 831-836.

[3] Belhumeur V, Hespanha J, Kriegman D. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans on Pattern Ananlysis and Machine Intelligence.* 1997, **19**(7): 711-720.

[4] J. Yang., J.Y.Yang. Why can LDA be performed in PCA transformed space?. *Pattern Recognition*. 2003, **36**(2): 563-566.

[5] Fubing CHEN, Shengliang ZHANG, Xiumei GAO, Jingyu YANG. Theory of Fisher Linear Discriminant Analysis for Small Sample size Problem and Its Verification. *Journal of Image and Graphics*. 2005, **10**(8): 984-991

[6] X.S.Zhuang, D.Q.Dai. Inverse Fisher Discriminant criteria for small sample size problem and its application to face recognition. *Pattern Recognition*. 2005, **38**(11): 2192-2194.

[7] X.S.Zhuang, D.Q.Dai. Improved discriminant analysis for high-dimensional data and its application to face recognition. *Pattern Recognition.* 2007, **40**(5): 1570-1578.

[8] M. Li, B. Yuan. 2D-LDA: a statistical linear discriminant analysis for image matrix. *Pattern Recognition Letters.* 2005, **26**(5): 527-532

[9] H. Xiong, M.N.Swamy, M.O. Ahmad. Two-dimensional FLD for face recognition. *Pattern Recognition.* 2005, **38**(7): 1121-1124.

[10] J. Yang, D.Zhang. Y.Xu, J.Y, Yang. Two-dimensional discriminant transform for face recognition. *Pattern Recognition.* 2005, **38**(7): 1125-1129

[11] YANG J, ZHANG D, YANG JY. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans on Pattern Ananlysis and Machine Intelligence*. 2004, **26**(1): 131-137.

[12] L.A.Zadeh. Fuzzy sets. *Info.Control.* 1965, **8**: 338-353

[13] Keller J M, Gray M R, Givern J. A. A fuzzy k-nearest neighbour algorithm. *IEEE Trans. Syst. Man Cybernet*. 1985, **15**(4): 580-585

[14] Kw K C, Pedry W. Face recognition using a fuzzy fisher classifier. *Pattern Recognition*. 2005, **38**(10): 1717-1732

[15] Yujie Zheng, Jingyu Yang et al. Fuzzy Kernel Fisher Discriminant algorithm with application to face recognition. *The 6th world Congress on Intelligent and Automation(WCICA06).* **12**(12): 9669-9672.

[16] Yujie Zheng, Jingyu Yang, Xiaojun Wu, YongZhi Li. A New Two-dimensional Linear Discriminant Analysis Algorithm Based on Fuzzy Set Theory. *Engineering Science.* 2007, **9**(2): 49-53

[17] Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience.* 1991, **3**(1): 71-86.

[18] P.J.Phillips, H. Moon, S.A.Rizvi, and P.J.Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence.* 2000, **22**(10): 1090-1104.

[19] P.J.Phillips. The Facial Recognition Technology (FERET) Database. http://www.itl.nist.gov/iad/humanid/feret/feret_master.html, 2004.