http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (4): 823-828, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.823.828

F-2DCCA: A New Fuzzy Feature Extraction Method for Face Recognition

[1]Yanyan Liu, [1]Yibin Zhao and [2]Jianguo Wang
[1]Department of Basic Courses, Institute of Disaster Prevention, Sanhe 065201, China
[2]Multimedia Making Center, Beijing Information Technology College, Beijing 100015, China

Abstract: Integrating the sample distribution information into the process of feature extraction is beneficial to classification accuracy. In this paper, a fuzzy two-dimensional canonical correlation analysis (F-2DCCA) method is proposed for image feature extraction. By making use of the Fourier transform and fuzzy *k*-nearest neighbor algorithm, we first construct a new fuzzy class-membership matrix to represent the distribution of image samples. Furthermore, two improvements based on Two-dimensional Canonical Correlation Analysis (2DCCA) are presented to promote the discrimination performance of the feature vectors and reduce their dimension respectively. The experimental results on the combined face database demonstrate the feasibility and effectiveness of the proposed approach.

Key words: Two-dimensional canonical correlation analysis, fuzzy *k*-nearest neighbor, Fourier spectrum, criterion function, face recognition

INTRODUCTION

As a successful application to identify person, face recognition has received increasing attention in computer vision and pattern recognition (Zhao *et al*., 2003; Jafri and Arabnia, 2009). Feature extraction is a crucial step for face recognition to reduce the data dimensionality and to enhance the discriminatory information. The most well-known methods for facial feature extraction are eigenfaces (Turk and Pentland, 1991) and Fisherfaces (Belhumeur *et al*., 1997), whose underlying ideas are Principal Component Analysis (PCA) and Fisher Linear Discriminant Analysis (FLDA) respectively. PCA is a typical case of unsupervised feature extraction, which is oriented toward the optimal representation in a low-dimensional subspace but ignores the class information totally. In contrast, FLDA as a supervised feature extraction method produces separated classes in a low-dimensional subspace by maximizing the ratio of between-class scatter to within-class scatter. Therefore, the feature extracted by FLDA is called the most discriminating feature (Swets and Weng, 1996; Pang *et al*., 2006).

Recently Canonical Correlation Analysis (CCA) gained much attention in the field of face recognition (He *et al*., 2005; Sun and Chen, 2007; Liu *et al*., 2008). CCA was initially proposed as a multivariate analysis method to measure the linear relationships between two sets of variables (Hotelling, 1936). When taking one set of variables to represent the class information, CCA can be used for supervised feature extraction (Sun and Chen, 2007; Liu *et al*., 2008). The class information usually appears in the form of numerical labels, especially if these labels are encoded as binary vectors, CCA is equivalent to FLDA (Barker and Rayens, 2003). Evidently, CCA may obtain potential classification performance by utilizing appropriate class labels which reflect the sample distribution. Unfortunately, the CCA-based methods always suffer from the singularity problem of covariance matrix in high-dimensional face image recognition application. In order to solve this problem, two-dimensional CCA (2DCCA) method (Zou *et al*., 2007; Sun *et al*., 2010) is developed by directly performing the image matrices to compute the image covariance matrix, which can effectively avoid the singularity problem and achieve better recognition accuracy. However, 2DCCA creates a binary class-membership matrix with 0 and 1 to represent the relationship between the samples and classes (Zou *et al*., 2007). Face images are actually affected by numerous environmental conditions and the boundaries between different classes may become indistinct (Sun and Chen, 2007). These binary class labels do not consider the samples located in or near the overlapping regions among classes (Liu *et al*., 2008). Therefore, 2DCCA method will hard to avoid losing the significant discriminatory information. Since the near classes overlap each other, an improved class-membership matrix considered the neighborhood information may be more suitable to reflect the sample distribution. Fuzzy set theory is hence a natural choice to

Corresponding Author: Yanyan Liu, Department of Basic Courses, Institute of Disaster Prevention, Sanhe, 065201, China
Tel: +8615910398179

yield fuzzy class labels by using membership degrees. In addition, for its number of columns is same as the image matrix (Sun *et al*., 2010), the feature matrix extracted by 2DCCA usually requires heavy storage space and process time.

According to the above discussion, we present a fuzzy 2DCCA (F-2DCCA) method in this paper, which overcomes the shortcomings of 2DCCA and reserves its advantage to extract combined features. By utilizing the fuzzy set theory, we first construct a new class-membership matrix based on Fourier spectrum representation of face images to reflect the distribution of samples. Subsequently, a modified criterion function is proposed to promote the recognition performance. Finally, two-dimensional PCA (2DPCA) method (Yang *et al*., 2004; Jing *et al*., 2006) is used to further reduce the feature dimension. Experimental results on face databases show that the proposed method is efficient and practical.

TWO-DIMENSIONAL CANONICAL CORRELATION ANALYSIS

Suppose there are c known image classes and the ith class has N_i samples. The jth training sample in class i is denoted by an m×n matrix A_{ij}. Based on the given image matrices, the centered sample matrix set can be constructed by:

$$X_A = (A_{11} - \bar{A}, \ldots, A_{1N_1} - \bar{A}, \ldots, A_{cN_c} - \bar{A}) \quad (1)$$

where, $\bar{A}$ is the total mean. The class label matrix corresponding to sample A_{ij} can be defined as:

$$B_{ij} = (B^1_{ij}, B^2_{ij}, \ldots, B^c_{ij})^T \in R^{(n\times c)\times m} \quad (2)$$

where, B^1_{ij} is an m×n matrix with ones on the diagonal and zeros elsewhere, $B^k_{ij}(k\neq i)$ is an m×n zero matrix and the symbol T denotes transpose. So, the centered class label matrix set, called class-membership matrix, can be denoted as:

$$Y_A = (B_{11} - \bar{B}, \ldots, B_{1N_1} - \bar{B}, \ldots, B_{cN_c} - \bar{B}) \quad (3)$$

Where:

$$\bar{B} = \frac{1}{N}\sum_{i,j} B_{ij}$$

and

$$N = \sum_{i=1}^{c} N_i$$

The goal of 2DCCA is equivalent to finding pairs of projection directions α and β that maximize:

$$J(\alpha, \beta) = \frac{\alpha^T X_A Y_B^T \beta}{\sqrt{\alpha^T X_A X_A^T \alpha \beta^T Y_B Y_B^T \beta}} \quad (4)$$

This can be expressed as the following generalized eigenvalue equation:

$$(X_A X_A^T)^{-1} X_A Y_B^T (Y_B Y_B^T)^{-1} Y_B X_A^T \alpha = \lambda\alpha \quad (5)$$

Generally, the image covariance matrix $X_A X_A^T$ is nonsingular when $N \geq c+1$, which is always satisfied in real image recognition task. As to the class-membership covariance matrix $Y_B Y_B^T$, we need to compute the analytical solution of generalized inverse for it is always singular. Solving Eq. 5, we can obtain eigenvectors α_i corresponding to the first d largest eigenvalues. Thereby the 2DCCA feature matrix of sample $A \in R^{m\times n}$ can be calculated as:

$$Z = (\alpha_1, \alpha_2, \ldots, \alpha_d)^T (A - \bar{A}) \in R^{m\times n} \quad (6)$$

Besides, Lee and Choi (2007) and Song *et al*. (2008) have proposed different 2DCCA algorithms for feature fusion. As we focus our efforts on supervised feature extraction, those algorithms will not be discussed in this study.

FUZZY TWO-DIMENSIONAL CANONICAL CORRELATION ANALYSIS

Fuzzy class-membership matrix: 2DCCA can be used to extract combined features by correlating samples with the sample distribution information. But how to represent the distribution of image samples in the form of numerical values appropriately is a key and difficult problem (Liu *et al*., 2008). As the environmental influence blurs the boundaries between classes, image samples may have relations with each class. These relationships can be represented by fuzzy membership degrees. In our method, Fuzzy *k*-nearest Neighbor (FKNN) method (Sun and Chen, 2007) is utilized to gain the fuzzy membership degree for each sample by making use of neighborhood information. More specifically, for training sample A_{ij} coming from class i, its corresponding fuzzy class label vector can be defined by:

$$W_{ij} = (u_{1j}, u_{2j}, \ldots, u_{cj}) \in R^c \quad (7)$$

Where:

$$u_{mj} = \begin{cases} 0.51 + 0.49 n_{mj}/k, & m = i \\ 0.49 n_{mj}/k & m \neq i \end{cases} \quad (8)$$

In the above expression, n_{mj} stands for the number of A_{ij}'s neighbors coming from class m and k is the parameter of *k*-nearest neighbor method. Obviously, the fuzzy class labels vector w_{ij} reflects the membership degree of A_{ij} belonging to each image class. In other words, all the w_{ij} ($i = 1, \ldots, c; j = 1, \ldots, N_i$) can describe the approximate information of sample distribution in some sense.

To adapt the 2D feature extraction algorithm, the fuzzy class label vectors w_{ij} should be disposed in an appropriate matrix form. Since the image covariance matrix $X_A X_A^T$ is essentially evaluated using the column vectors of the sample matrices (Kong *et al*., 2005), we can utilize the class label vector w_{ij} for each column vector of A_{ij}, respectively. Owning to its translation invariance (Lai *et al*., 2001), the Fourier transform is applied to the face images, which can align the column vectors and eliminate the effect of displacement. Let $j = \sqrt{-1}$ denotes the imaginary unit, for an m×n face image A, 2D Fourier spectrogram is defined as:

$$F = |F(u, v)| = [R^2(u, v)+I^2(u, v)]^{1/2} \quad (9)$$

Where:

$$F(u,v) = \frac{1}{mn}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} A(x,y)e^{-j2p(ux/m+vy/n)} \quad (10)$$

Hence the centered sample spectrogram set can be constructed by:

$$X_F = (F_{11} - \bar{F},\ldots, F_{1N_1} - \bar{F},\ldots, F_{cN_c} - \bar{F}) \quad (11)$$

Where:

$$\bar{F} = \frac{1}{N}\sum_{i,j} F_{ij}$$

Furthermore, Nastar and Ayache (1996) found that facial expressions and small occlusion only affect the high-frequency spectrum and Bow (1992) concluded that the image can be reconstructed by the lower half-frequency spectrum. The above statements show that the column vector contained more low-frequency information play a more dominant role in face recognition. As the low-frequency component usually has a bigger value, we defined the fuzzy class label matrix corresponding to sample A_{ij} as:

$$E_{ij} = ((\gamma_{ij})_1 w_{ij},\ldots, (\gamma_{ij})_2) w_{ij},\ldots, (\gamma_{ij})_n) w_{ij})\epsilon R^{c\times n} \quad (12)$$

Where:

$$(\gamma_{ij})_d = \frac{\text{sum}[(F_{ij})_d]}{\max\limits_{s=1,\ldots,n}\left(\text{sum}\left[(F_{ij})_s\right]\right)} \in (0,1] \quad (13)$$

In Eq. 13, $(F_{ij})_d$ is the dth column vector of spectrogram F_{ij} and $\text{sum}[(F_{ij})_d]$ denotes the sum of all the elements in $(F_{ij})_d$. The value of $(\gamma_{ij})_d$ will be bigger if the vector $(F_{ij})_d$ contain more low-frequency information or vice versa. That is to say, $(\gamma_{ij})_d$ exhibits the weightiness of $(F_{ij})_d$ in feature extraction. Thus the centered fuzzy class label matrix set, called fuzzy class-membership matrix can be denoted as:

$$Y_E = (E_{11} - \bar{E},\ldots, E_{1N_1} - \bar{E},\ldots, E_{cN_c} - \bar{E}) \quad (14)$$

where, $\bar{E}$ is the average matrix of all the fuzzy class label matrices.

F-2DCCA based feature extraction: Motivated by the Fisher criterion (Belhumeur *et al*., 1997) which aims to maximize the class separability, we modified the criterion function of F-2DCCA as follows:

$$\hat{J}(\alpha,\beta) = \frac{\alpha^T X_F Y_E{}^T \beta}{\sqrt{\alpha^T S_{WF}\alpha\beta^T Y_E Y_E{}^T\beta}} \quad (15)$$

where, S_{WF} is the within-class scatter matrix in the sample space X_F. Maximizing $\hat{J}(\alpha, \beta)$ aims to find pairs of projection directions, which can minimize the within-class scatter of the sample projections and maximize the correlation between X_F and Y_E simultaneity. Consequently, the features extracted by maximizing the function $\hat{J}(\alpha, \beta)$ can provide more discriminatory information. Similar to the derivations of 2DCCA, F-2DCCA is equivalent to solving the following generalized eigenvalue equation:

$$(S_{WF})^{-1}X_F Y_E{}^T(Y_E Y_E{}^T)^{-1} Y_E X_F{}^T\alpha = \lambda\alpha \quad (16)$$

Let α_i denote the eigenvectors corresponding to the first d largest eigenvalues. Then the feature matrix of sample A, whose 2D Fourier spectrogram is F, can be calculated as:

$$\hat{Z} = (\alpha_1,\ldots, \alpha_d)^T(F-\bar{F})\epsilon R^{d\times n} \quad (17)$$

Because the column number n of $\hat{Z}$ is still same as the image matrix, the feature matrix may consume much storage and time in the matching process. 2DPCA method (Yang *et al*., 2004) is chosen to reduce the feature matrix's dimension, inasmuch as it can greatly reduce the dimension and reserve the information in data as much as possible. Furthermore, an automatic strategy proposed by Jing *et al.* (2006) is adopted to select discriminative vectors and 2D principal components. This strategy can select h helpful eigenvectors η_i of 2DPCA automatically, where h is much smaller than n and these eigenvectors

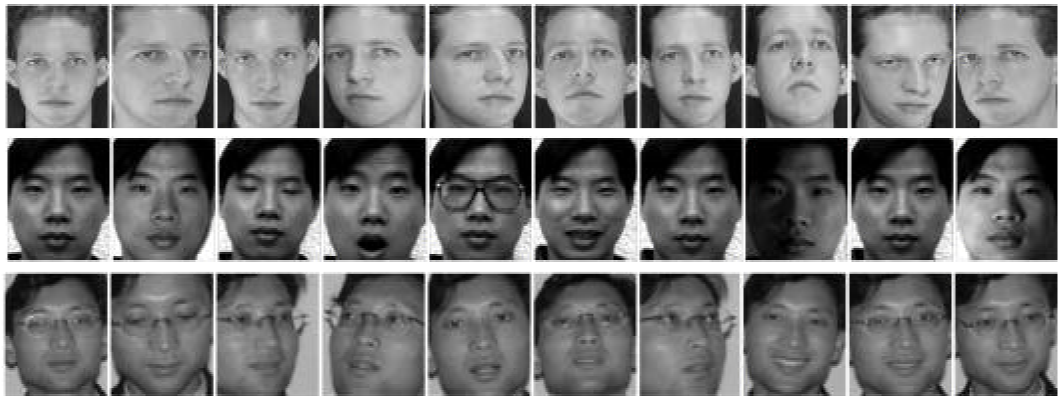

Fig. 1: Some face images from the combined database

contain more between-class separable information. Accordingly, the F-2DCCA features of a sample are calculated as:

$$\tilde{Z} = (\alpha_1,\ldots,\alpha_d)^T(F-\bar{F})(\eta_1,\ldots,\eta_h)\in R^{d\times h} \quad (18)$$

We anticipate that these low-dimension features $\tilde{Z}$ extracted by F-2DCCA have more powerful ability of classification, for they not only incorporate the information about spectrum and distribution of samples, but also try to minimize the within-class scatter and maximize the between-class separation.

EXPERIMENTAL RESULTS

To evaluate the performance of our method, some experiments are performed on a large face database by combining ORL, Yale and Lab database. Lab database contains 450 images from 30 members in our laboratory and these images are taken with little restriction. The combined database is composed of 85 distinct subjects. Each subject has 10 selected images, which are cropped into 112×92 and contain pose, illumination and expression variations. Figure 1 shows some images in the combined database. The first 10 largest eigenvectors α_i are used to extract features for classification by the nearest neighbor classifier and all experiments are programmed in MATLAB 7.1 environment.

First of all, we make experiments to compare F-2DCCA with 2DCCA method (Sun *et al*., 2010). We randomly choose 10 different training and testing sets from the combined database. The number of training samples per subject is 4. To reduce the computation complexity of 2DCCA, each face image in database is down-sampled to 28×23. The recognition results in each

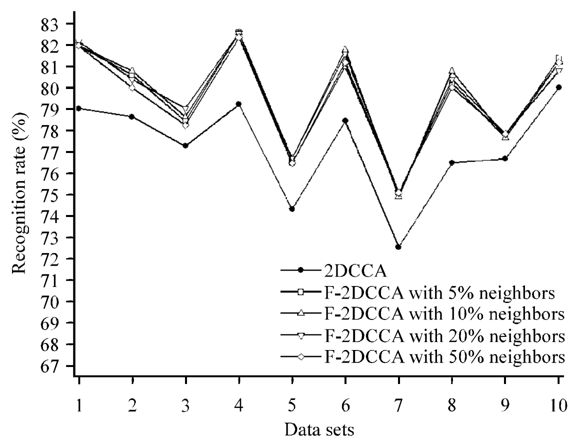

Fig. 2: Recognition rates comparison on 10 different data sets

round are shown in Fig. 2. Meanwhile, in order to show how the parameter k of FKNN affects F-2DCCA algorithm's performance, we change it by concerning different percentage of nearest neighbors, as shown in Fig. 2. From the experimental results, we can observe that the proposed method can achieve better accuracy than 2DCCA and the slight fluctuation of the recognition accuracy w.r.t. the parameter k reflects its robustness in some degree.

We then fix the parameter k by 10% neighbors for F-2DCCA and compare it with three one-dimensional (1D) methods, such as Fisherfaces (Belhumeur *et al*., 1997), PCA+CCA (He *et al*., 2005) and Fuzzy label CCA (Liu *et al*., 2008). The number of training samples per subject increases from 2 to 9. Training samples are selected randomly and the remaining samples are used for testing. Ten times of random selections are performed on

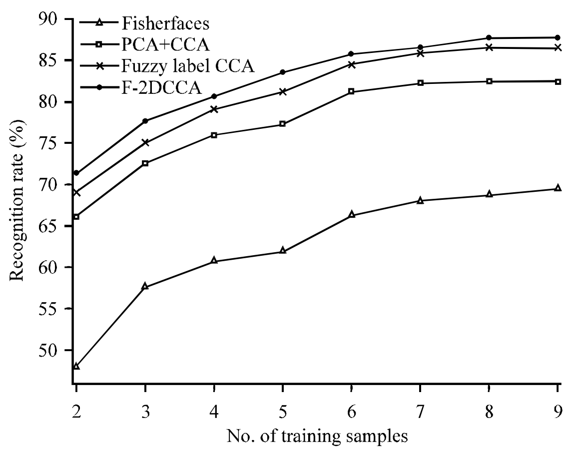

Fig. 3: Recognition rates comparison on the combined database

the combined database. In these experiments, the regularization parameter μ is taken as an experiential value 10^{-6} in the last two methods. Figure 3 shows the average recognition rates. It can be seen that F-2DCCA achieves better results than the other 1D methods. These results show that the sample distribution information and matrix-based image representation are both beneficial to classification.

CONCLUSION

In this study, a feature extraction method called F-2DCCA is proposed for face recognition. On one hand, F-2DCCA takes advantages of existing 2DCCA method to extract combined features integrating class information. On the other hand, F-2DCCA overcomes the shortcomings of 2DCCA method. Specifically, a new fuzzy class-membership matrix is defined and used in F-2DCCA to reflect the sample distribution, which is more reasonable and applicable than the binary one. Moreover, the features of F-2DCCA are more discriminative and lower dimensional than 2DCCA, which improve the computational efficiency. Experimental results show that the new approach is effective and feasible. Apparently, incorporating the sample distribution information into feature extraction is beneficial to classification. So how to construct more appropriate fuzzy class-membership matrix deserves further investigation.

ACKNOWLEDGMENTS

This study was supported by Teachers' Scientific Research Fund of China Earthquake Administration (No. 20110116) and Hebei Province Natural Sciences Foundation (No. A2011408006). The authors would like to thank the anonymous reviewers for their precious suggestions.

REFERENCES

Barker, M. and W. Rayens, 2003. Partial least squares for discrimination. J. Chemom., 17: 166-173.

Belhumeur, P.N., J.P. Hespanha and D.J. Kriegman, 1997. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intel., 19: 711-720.

Bow, S.T., 1992. Pattern Recognition and image Preprocessing. Marcel Dekker, New York.

He, Y., L. Zhao and C. Zou, 2005. Face recognition based on PCA/KPCA plus CCA. Proceedings of the International Conference on Natural Computation, August 27-29, 2005, Changsha, China, pp: 71-74.

Hotelling, H., 1936. Relations between two sets of variates. Biometrika, 28: 321-377.

Jafri, R. and H.R. Arabnia, 2009. A survey of face recognition techniques. J. Inform. Process. Syst., 5: 41-68.

Jing, X.Y., H.S. Wong and D. Zhang, 2006. Face recognition based on 2D Fisherface approach. Pattern Recognition, 39: 707-710.

Kong, H., L. Wang, E.K. Teoh, X.C. Li, J.G. Wang and R. Venkateswarlu, 2005. Generalized 2D principal component analysis for face image representation and recognition. Neural Networks, 18: 585-594.

Lai, J.H., P.C. Yuen and G.C. Feng, 2001. Face recognition using holistic Fourier invariant features. Patter Recognition, 34: 95-109.

Lee, S.H. and S. Choi, 2007. Two-dimensional canonical correlation analysis. IEEE Signal Process. Lett., 14: 735-738.

Liu, Y., X. Liu and Z. Su, 2008. A new fuzzy approach for handling class labels in canonical correlation analysis. Neurocomputing, 71: 1735-1740.

Nastar, C. and N. Ayache, 1996. Frequency-based nonrigid motion analysis: Application to four dimensional medical images. IEEE Trans. Pattern Anal. Mach. Intell., 18: 1067-1079.

Pang, Y.W., Z.K. Liu and N.H. Yu, 2006. A new nonlinear feature extraction method for face recognition. Neurocomputing, 69: 949-953.

Song, D.X., Y.J. Liu and C.K. Chen, 2008. Two-dimensional canonical correlation analysis and its application to face recognition. Comput. Appl., 28: 2430-2433.

Sun, N., Z.H. Ji, C.R. Zou and L. Zhao, 2010. Two-dimensional canonical correlation analysis and its application in small sample size face recognition. Neural Comput. Appl., 19: 377-382.

Sun, T. and S. Chen, 2007. Class label versus sample label-based CCA. Applied Math. Comput., 185: 272-283.

Swets, D.L. and J. Weng, 1996. Using discriminant eigenfeatures for image retrieval. IEEE Trans. Patt. Anal. Mach. Intel., 18: 831-836.

Turk, M. and A. Pentland, 1991. Eigenfaces for recognition. J. Cognitive Neurosci., 3: 71-86.

Yang, J., D. Zhang, A.F. Frangi and J. Yang, 2004. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell., 26: 131-137.

Zhao W., R. Chellappa, P.J. Phillips and A. Rosenfeld, 2003. Face recognition: A literature survey. ACM Comput. Surveys, 35: 399-458.

Zou, C.R., N. Sun, Z.H. Ji and L. Zhao, 2007. 2DCCA: A novel method for small sample size face recognition. Proceedings of the 8th IEEE Workshop on Applications of Computer Vision, February 21-22, 2007, Austin, Texas, USA, pp: 43-43.