

Fuzzy Maximum Scatter Discriminant Analysis and Its Application to Face Recognition

Jianguo Wang^{1,2} Wankou Yang¹ Jingyu Yang¹

¹ School of Computer Science & Technology, Nanjing University of Science & Technology, Nanjing, PR China

² Network & Education Center, Tangshan College, Tangshan, PR China

wjgfwjg@163.com, wankou_yang@yhoo.com.cn, yangjy@mail.njust.edu.cn

Abstract

In this paper, a reformative scatter difference discriminant criterion (SDDC) with fuzzy set theory is studied. The scatter difference between between-class and within-class as discriminant criterion is effective to overcome the singularity problem of the within-class scatter matrix due to small sample size problem occurred in classical Fisher discriminant analysis. However, the conventional SDDC assumes the same level of relevance of each sample to the corresponding class. So, a fuzzy maximum scatter difference analysis (FMSDA) algorithm is proposed, in which the fuzzy k-nearest neighbor (FKNN) is implemented to achieve the distribution information of original samples, and this information is utilized to redefine corresponding scatter matrices which are different to the conventional SDDC and effective to extract discriminative features from overlapping (outlier) samples. Experiments conducted on FERET face databases demonstrate the effectiveness of the proposed method.

1. Introduction

It is well known that Fisher linear discriminant analysis (FLDA) is a popular and effective method for feature extraction. The basic idea of FLDA is try to find an optimal projection vectors by maximizing between-class scatter matrix and minimizing within-class scatter matrix.

However, due to face recognition frequently encountered the high dimensionality and small sample size [1,2], the classical FLDA cannot be used directly in that the within-class scatter matrix is always singular. To overcome this problem, some

corresponding techniques were proposed [1,3,4,5]. The most popular method, called Fisherface, was build by Swets [6] and Belhumeur [7]. 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 $d (d \leq c)$. Obviously, in the process K-L transform, the small 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 be not singular.

To avoid the singularity problem, Song et al. [8] proposed a method, which adopts the difference of both between-class scatter and within-class scatter as discriminant criterion, due to the inverse matrix is need not constructed, so the small sample size problem occurred in traditional Fisher discriminant analysis is in nature avoided.

However, this method was employed to dwell on binary class assignment meaning that the samples come fully assigned to the given classes, and in the real world, samples are always affected by environmental conditions (face images are affected by illumination, pose, etc.). To address this problem, we reexamine the scatter difference discriminant method and introduce fuzzy set theory to augment it. By taking advantage of the technology of fuzzy set theory, a new feature extraction method named fuzzy maximal scatter difference method is proposed in this paper. Experimental results on the FERET face database demonstrate the effectiveness of the proposed method.

The paper of the rest is organized as follows. In section 2, we review briefly maximum scatter difference criterion. In section 3, we propose the idea and describe the proposed method in detail. In section 4, experiment on FERET face database is presented to

demonstrate the effectiveness of the proposed method. Conclusions are summarized in section 5.

2. Scatter difference discriminant criterion

Suppose there are C known pattern classes, $\omega_1, \omega_2, \dots, \omega_C$, the between-class scatter matrix and within-class scatter matrix can be denoted as:

$$S_b = \frac{1}{M} \sum_{i=1}^C M_i (m_i - m_0)(m_i - m_0)^T \quad (1)$$

$$S_w = \frac{1}{M} \sum_{i=1}^C \sum_{j=1}^{M_i} (x_i^j - m_i)(x_i^j - m_i)^T \quad (2)$$

Where M is the total number of training samples, and M_i is the number of training samples in class i .

In class i , the j -th training sample is denoted by x_i^j , the mean vector of training samples in class i is denoted by m_i and the mean vector of all training samples is m_0 .

From the classical Fisher criterion function [3], we know that when the ratio of the between-class scatter and the within-class scatter is maximized, the samples can be separated easily. In this paper, a scatter difference based discriminant rule is defined as follows [8]:

$$\begin{aligned} J_d(w) &= w^T S_b w - B \cdot w^T S_w w \\ &= w^T (S_b - B \cdot S_w) w \end{aligned} \quad (3)$$

Where B is a nonnegative constant to balance S_b and S_w . By the property of the extreme value of generalized Rayleigh quotient [9], the optimal solution set maximizing (3) are the eigenvectors w_1, w_2, \dots, w_k corresponding to the first k largest eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$, i.e., $(S_b - B \cdot S_w)w_j = \lambda_j w_j$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$.

Comparing the Maximum Scatter Difference (MSD) criterion with the classical Fisher discriminant criterion, we find that the former avoids calculation of the inverse within-class scatter, i.e., $S_w^{-1} S_b$ is substituted by $S_b - B \cdot S_w$, this can not only make computationally more efficient but also avoid the singular problem of the within-class scatter.

3. Fuzzy maximum scatter difference (FMSD) criterion

3.1 Fuzzy k-nearest neighbor algorithm

Fuzzy set theory is the generalization of the classical set theory, and fuzzy pattern recognition is that the fuzzy logic method is introduced to solve the classical pattern recognition problem.

In this paper, FKNN algorithm [10] is adopted to calculate fuzzy membership degree and then which is used to calculate each class center. Suppose there are C known pattern classes, and M training samples, let $U = [\mu_{ij}]$, $i = 1, 2, \dots, C$, $j = 1, 2, \dots, M$, where μ_{ij} indicates the degree that the j th sample belongs to class i . Thus with the FKNN algorithm, the following provides the algorithm of computations of the fuzzy membership degrees (μ_{ij}):

Step 1. Compute the Euclidean distance matrix between any pair of feature vectors in the training samples set.

Step 2. Set diagonal elements (Euclidean distances of samples and themselves, which equal to zeroes) of the Euclidean distance matrix to infinity.

Step 3. 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 k-neighborhood.

Step 4. Compute the membership degree of the samples using Eq. (4).

$$\mu_{ij} = \begin{cases} 0.51 + 0.49(n_{ij} / k), & \text{if the } j\text{th sample} \\ & \text{belongs to class } i \\ 0.49(n_{ij} / k), & \text{otherwise} \end{cases} \quad (4)$$

Where n_{ij} stands for the number of the neighbors of the j th pattern that belong to the class i .

After obtain the fuzzy membership degrees (μ_{ij}), and then we can adjust each class center using Eq. (5).

$$\omega_i = \frac{\sum_{j=1}^M \mu_{ij} x_j}{\sum_{j=1}^M \mu_{ij}} \quad (5)$$

Thus we also get the corresponding class center matrix ω , $\omega = [\omega_i]$, $i = 1, 2, \dots, C$.

3.2 The Proposed Method (FMSD)

How to make full use of distribution information of the training samples is the key step of our method. We first calculated the fuzzy membership degree and the new center of each class with FKNN algorithm, and redefined the samples' scatter matrices, then, we can use these redefined scatter matrices to design an MSD algorithm, called FMSD.

In the redefinition of the between-class scatter matrix, we considered that a class which is far from the total center will have more contribution to classification, while in the redefinition of the fuzzy within-class scatter matrix, samples that are more close to the class center have more contributions to classification, so the membership degree of each sample are considered in the redefinition of the corresponding fuzzy scatter matrices, and the fuzzy between-class scatter matrix, the fuzzy within-class scatter matrix and the fuzzy total scatter matrix are defined as follows.

$$FS_b = \sum_{i=1}^C \left[\left(1 - \sum_{y_j \in \omega_i} \mu_{ij}^p / \sum_{j=1}^M \mu_{ij}^p \right) (\omega_i - \bar{y})(\omega_i - \bar{y})^T \right] \quad (6)$$

$$FS_w = \sum_{i=1}^C \left[\sum_{y_j \in \omega_i} \mu_{ij}^p (y_j - \omega_i)(y_j - \omega_i)^T \right] \quad (7)$$

Where \bar{y} is the mean of all training samples, and p is a constant value that determines how heavily the fuzzy membership degree is weighted when calculating each scatter matrix.

From the aforementioned equations (3), (6) and (7), we can obtain the fuzzy scatter difference discriminant criterion described as Eq. (8).

$$J_{Fd}(w) = w^T FS_b w - B \cdot w^T FS_w w \\ = w^T (FS_b - B \cdot FS_w) w \quad (8)$$

Then the optimal discriminant eigenvectors matrix can be calculated with the eigensystem.

$$(FS_b - B \cdot FS_w) w = \lambda w \quad (9)$$

After obtaining the projection axes, we can form the following linear transform for a given sample x_i .

$$f_i = W_F^T \cdot x_i \quad (10)$$

The feature vector f_i is used to represent the sample x_i for recognition purposes.

In summary of the preceding description, the proposed algorithm is described as follows.

Step 1. Calculate the fuzzy membership degree matrix U and the new class center matrix ω of training samples using Eqs. (4), and (5) respectively.

Step 2. Calculate the fuzzy between-class scatter matrix FS_b , and fuzzy within-class scatter matrix FS_w using Eqs. (6), and (7) respectively.

Step 3. Calculate the transformation matrix W_F of the proposed algorithm using Eq. (9), and project all the samples onto the W_F , then obtain the projection coefficients (feature vectors) $F = W_F^T \cdot X$.

Step 4. Classify the projection coefficients.

4. Experiment on the FERET database

The proposed method was tested on a subset of the FERET database. This subset includes 1,400 images of 200 persons (each person has seven images), which involves variations in facial expression, illumination, and pose. In our experiment, the facial portion of each original image is cropped automatically based on the location of eyes and resized to 40*40 pixels and without histogram equalization. Some facial portion images of one person are shown in Fig.1.



Figure 1. Sample images of one person in FERET database

To evaluate the recognition performance of the proposed method, we run the system 7 times. In each time, three images per class are selected for training and the rest four images are used for testing, thus the training samples number is 600 and the testing samples number is 800. In the i -th test, the training samples are i , $i+1$, $i+2$ per class, respectively. (When $i+1$ and $i+2$ are larger than 7, then substitute $i+1-7$ and $i+2-7$ for them). In the experiment, PCA, LDA, MSD, Fuzzy LDA, and the proposed method (Fuzzy MSD) are used for feature extraction. Finally, the nearest neighbor (NN) classifiers and the minimal distance (MD) classifiers with Euclidean distance and cosine distance are, respectively, employed for classification. The average recognition rates and standard deviations (std) across 7 runs of each method and their corresponding dimension are shown in Table 1.

From table 1, we can see that Fuzzy method improves the performance of the MSD method markedly under the minimal distance classifiers with Euclidean distance. While under the nearest neighbor classifier, no matter what distance (Euclidean distance and Cosine distance) is adopted, the performance of the Fuzzy MSD is not improved on the whole. Why

Table 1. The mean and standard deviation of recognition rates (%) of PCA, LDA, MSD, Fuzzy LDA, and the proposed method (Fuzzy MSD) on the FERET database under the nearest neighbor (NN) classifiers and the minimal distance (MD) classifiers with Cosine distance and Euclidean distance when the three samples (i , $i+1$, $i+2$) per class are used for training

Methods	Euclidean		Cosine	
	MD	NN	MD	NN
PCA	42.04 ± 9.96	52.75 ± 7.94	40.3 ± 8.74	47.93 ± 8.32
LDA	49.57 ± 11.45	44.29 ± 10.81	51.05 ± 10.75	50.2 ± 10.41
MSD	49.52 ± 10.22	55.79 ± 8.63	57.86 ± 10.57	57.39 ± 11.48
Fuzzy LDA	51.45 ± 9.8	45.07 ± 8.58	53.66 ± 9.45	51.48 ± 9.74
Fuzzy MSD	57.91 ± 8.54	55.95 ± 8.11	59.63 ± 9.67	57.89 ± 9.25

this happens? As we know, the fuzzy method changes the distribution of the samples, but which does not change the class of the samples. When classification with the minimal distance classifier, we calculate the distance between test sample and the fuzzy class centers, so the performance is improved.

5. Conclusion

In this paper, a new subspace method, fuzzy maximal scatter difference (FMSD) discriminant criterion, is proposed to extract features from samples, especially from overlapping (outlier) samples. This algorithm is based on conventional maximal scatter difference method, which can avoid the singularity problem of the within-class scatter matrix due to small sample size problem occurred in classical Fisher discriminant analysis. Furthermore, the overlapping (outlier) samples' distribution information is incorporated in the redefinition of corresponding scatter matrices, which is important for classification. The experiments are conducted on a subset of FERET database with several feature extraction methods and the proposed method; the experiment results indicate that the proposed method has better recognition rate, it is shown that the accuracy can be further improved by combining Fuzzy set theory and MSD together.

Acknowledgements

This work was supported by the following projects: National Natural Science Foundation China (Grant No. 60632050, 60503026 and 60472060), the Graduates' Research Innovation Program of Higher Education of Jiangsu Province (Grant No. CX07B_118z).

References

- [1] L.F. Chen, H.Y.M. Liao, M.T. Ko, et al. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 2000, 33 (10): 1713~1726.
- [2] H. Yu, J. Yang. A Direct LDA Algorithm for High-Dimensional Data With Application to Face Recognition. *Pattern Recognition*, 2001, 34(10): 2067-2070.
- [3] W. Zhao, R. Chellappa, J. Phillips. Subspace Linear Discriminant Analysis for Face Recognition. Technical Report CS-TR4009, Univ. of Maryland, 1999.
- [4] J. Yang, J.Y. Yang. Why can LDA be performed in PCA transformed space?. *Pattern Recognition*. 2003,36(2): 563-566.
- [5] 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.
- [6] D. L. Swets, J. Weng. Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1996, 18(8): 831-836.
- [7] V. Belhumeur, J. Hespanha, D. Kriegman. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans on Pattern Analysis and Machine Intelligence*. 1997, 19(7): 711-720.
- [8] F. X. Song, K. Cheng, J. Y. Yang, et al.. Maximum scatter difference, large margin linear projection and support vector machines. *Acta automatica sinica*, 2004, 30(6): 890-896. (in Chinese).
- [9] Zhang Xian-da. Matrix analysis and application (in Chinese). Beijing: Tsinghua University Press, 2004.
- [10] J. M. Keller, M. R. Gray, J. R. Givens. A fuzzy k-nearest neighbor algorithm. *IEEE Trans Syst Man Cybernet*, 1985,15(4): 580~585.