

Feature Extraction Using Fuzzy Complete Linear Discriminant Analysis

Yan Cui

School of Computer Science and Technology,
Nanjing University of Science and Technology,
Nanjing, 210094, P. R. China
Email: cuiyan899@163.com

Zhong Jin

School of Computer Science and Technology,
Nanjing University of Science and Technology,
Nanjing, 210094, P. R. China
Email: zhongjin@njust.edu.cn

Abstract—In pattern recognition, feature extraction techniques are widely employed to dimensionality reduction. In this paper, a novel feature extraction method, fuzzy complete linear discriminant analysis (Fuzzy-CLDA), is proposed by combining the complete linear discriminant analysis (CLDA) and the membership degrees of samples. Furthermore, we calculate the sample membership degrees with different distance metrics and compare the effectiveness of the distance metrics. In addition, experiments are provided for analyzing and illustrating our results.

I. INTRODUCTION

In statistical pattern recognition, particularly in image recognition, the original data are always in high-dimensional space and the classification in such a space is usually a difficult task due to some problems that characterize high-dimensional space known as "dimensionality curse" [1]. So, feature extraction techniques are widely employed to dimensionality reduction and to extract the most significant features. Principal component analysis (PCA) is a typical feature extraction method [2]. Fisher linear discriminant analysis (LDA) is another popular linear dimensionality reduction method [3]. However, LDA cannot be applied when the scatter matrix S_w or S_t is singular due to the small sample size problems. In the past, many LDA extensions have been developed to deal with this problem. Briefly, some extensions as: pseudo-inverse LDA (PLDA) [8], regular LDA (RLDA) [9], penalized discriminant analysis (PDA) [10], LDA/GSVD [11], LDA/QR [12], orthogonal LDA (OLDA) [13], null space LDA (NLDA) [14], direct-LDA (D-LDA) [15], CLDA [16] and two-stage LDA [17].

It is important to realize that the conventional optimality criterion defined based on the scatter matrices is not directly related to the samples distribution information. Specially, for multi-class classification problems involving more than two class, it overemphasizes the classes with large distance in the input space and causes large overlaps of neighbor classes in the output space [18]. To tackle this problem, a weighted function is incorporated into the Fisher criterion by giving higher weights to classes that are closer together in the input space as they are likely to lead to misclassification. Therefore the samples distribution information is represented by the weighted function. Recently, The fractional-step linear discriminant analysis (FLDA) in [19] allows for the relevant

distances to be more weighted. The direct fractional linear discriminant analysis (DFLDA) in [20] attempts to handle high-dimensional face patterns by combining FLDA and D-LDA algorithms. The weighted pairwise Fisher LDA in [18] incorporates a weighted function to the pairwise Fisher LDA which decomposes k -class Fisher LDA to $\frac{1}{2}k(k-1)$ two-class Fisher LDA. The uncorrelated weighted linear discriminant analysis (UWLDA) in [21] integrates the weighted scheme and searches the significant information in and out of the null space of S_w under the statistically uncorrelated constraints, respectively. Due to the weighted function is hard to determine, a number of studies based on the fuzzy set theory have been carried out for fuzzy pattern recognition, the samples distribution information is represented by the membership degrees corresponding to every class. The fuzzy Fisherface in [22] integrates the sample membership degree to the scatter matrix to improve classification results. The fuzzy inverse Fisher discriminant analysis (FIFDA) in [23] extracts discriminative information by combining the inverse Fisher discriminant criterion and the membership degree.

Inspired by ideas in [16, 22, 23], in this paper, we will propose a new feature extraction criterion, Fuzzy-CLDA, by combining CLDA and the sample membership degrees. Fuzzy-CLDA will extract discriminant vectors in and out of the null space of S_{wF} , respectively. The membership degree matrix in [22, 23] is computed by the Euclidean distance, but in Fuzzy-CLDA, we use 7 kinds of distance metrics to compute the membership degree matrix and obtain the optimal distance metric by cross-validation. The organization of this paper is as follows. We review briefly CLDA in section II. We will propose the Fuzzy-CLDA and describe the new method in detail in section III. In section IV, experiments are presented to demonstrate the effectiveness of the distance metric and Fuzzy-CLDA. Conclusions are summarized in section V.

II. THE COMPLETE LINEAR DISCRIMINANT ANALYSIS

In this section, we give an overview of CLDA, which is used to extract the discriminant vectors in the null and range space of S_w . For the given pattern samples data set $T = \{(x_1, y_1), \dots, (x_n, y_n)\} \in X \times Y$, $X = [x_1, \dots, x_n] \in R^{m \times n}$ is the input data matrix and $Y = \{C_1, \dots, C_c\}$ is the class label set. Let X be partitioned to c classes as

$X = [X_1, \dots, X_c]$, where $X_i \in R^{m \times n_i}$ and $\sum_{i=1}^c n_i = n$, N_i the the set of column indices that belong to the i -th class, that is, x_j , for $j \in N_i$, belongs to the i -th class. The within-class, between-class and total scatter matrix are defined as follows:

$$\begin{aligned} S_w &= \sum_{i=1}^c \sum_{j \in N_i} (x_j - m_i)(x_j - m_i)^T, \\ S_b &= \sum_{i=1}^c \sum_{j \in N_i} (m_i - m)(m_i - m)^T \\ &= \sum_{i=1}^c n_i (m_i - m)(m_i - m)^T, \\ S_t &= \sum_{i=1}^n (x_i - m)(x_i - m)^T, \end{aligned}$$

Where $m_i = \frac{1}{n_i} \sum_{j \in N_i} x_j$ is the mean vectors of the class C_i . Moreover, it can be verified that $S_t = S_w + S_b$ (see [11]). Since S_w, S_b and S_t are all semi-positive definite, it has:

Lemma 1. If S_t is singular, then $x^T S_t x = 0$ if and only if $x^T S_w x = 0$ and $x^T S_b x = 0$.

In this paper, we do not give the detailed derivation, thus some interested reader can be found the proof in [17]. According to Lemma 1, we can conclude that the null space of S_t contains any discriminative information, and hence we project the data onto the range space of S_t . We recompute the scatter matrices \hat{S}_w and \hat{S}_b in the reduced space. For the small sample size problems, the scatter matrix \hat{S}_w may be still singular, so we discuss the two case:

Case 1. If \hat{S}_w is nonsingular, we apply the Fisher LDA in the reduced space, i.e., solving the trace optimizations as follow:

$$J(G) = \max_G \text{trace}(G^T \hat{S}_w G)^{-1} (G^T \hat{S}_b G),$$

Where G is the linear transformation in the reduced space, $\hat{S}_w = P^T S_w P$, $\hat{S}_b = P^T S_b P$ and P is the orthogonal bases of the range space of S_t .

Case 2. If \hat{S}_w is singular, we extract irregular discriminant vectors in the null space $N(\hat{S}_w)$, and regular discriminant vectors in the range space $R(\hat{S}_w)$.

Firstly, We calculate the optimal irregular discriminant vectors. For an arbitrary vector $w \in N(\hat{S}_w)$, the Fisher criterion will definitely reach infinite. Therefore, $N(\hat{S}_w)$ is important to extract discriminant vectors. In this scenario, it is reasonable to use the between-class scatter matrix \hat{S}_b to measure the discriminative ability of a projection axis [14]. Thus the criterion degenerates into

$$J_b(w) = w^T \hat{S}_b w \quad (w^T w = 1). \quad (1)$$

Suppose ϕ_1, \dots, ϕ_r are the orthogonal eigenvectors of \hat{S}_w corresponding to zero eigenvalues, thus the null space $N(\hat{S}_w)$ can be denoted as $N(\hat{S}_w) = \text{span}\{\phi_1, \dots, \phi_r\}$. According to linear algebra, $N(\hat{S}_w)$ is isomorphic to r -dimensional Euclidean space R^r . The corresponding isomorphic mapping is

$$w = P_1 v, \text{ where } P_1 = (\phi_1, \dots, \phi_r), w \in N(\hat{S}_w), v \in R^r, \quad (2)$$

Which is a one-to-one mapping from R^r onto $N(\hat{S}_w)$. With this mapping, the criterion $J_b(w)$ in Eq.(1) is turned into

$$J_b(v) = v^T P_1^T \hat{S}_b P_1 v = v^T \tilde{S}_b v \quad (v^T v = 1), \quad (3)$$

Where $\tilde{S}_b = P_1^T \hat{S}_b P_1$. Note that \tilde{S}_b is $r \times r$ positive definite matrix, thus we can only generate r optimal vectors v_1, \dots, v_r . Then we can obtain $w_j = P_1 v_j$ ($j = 1, \dots, r$) by Eq.(2). From the property of isomorphic mapping, it is easy to know that $W_{ir} = [w_1, \dots, w_r]$ are the optimal irregular discriminant vectors with respect to $J_b(w)$.

Hereafter, we can calculate the optimal regular discriminant vectors w_{r+1}, \dots, w_{r+s} in a similar way. Suppose $\phi_{r+1}, \dots, \phi_{r+s}$ are the orthogonal eigenvectors of \hat{S}_w corresponding to the nonzero eigenvalues, thus the $N(\hat{S}_w)^\perp$ space can be denoted as $N(\hat{S}_w)^\perp = \text{span}\{\phi_{r+1}, \dots, \phi_{r+s}\}$. According to linear algebra, $N(\hat{S}_w)^\perp$ is isomorphic to s -dimensional Euclidean space R^s . The corresponding isomorphic mapping is

$$w = P_2 v, \text{ where } P_2 = (\phi_{r+1}, \dots, \phi_{r+s}), w \in N(\hat{S}_w)^\perp, v \in R^s. \quad (4)$$

With this mapping, the Fisher criterion is turned into

$$J(v) = \frac{v^T \bar{S}_b v}{v^T \bar{S}_w v}, \quad (5)$$

Where $\bar{S}_b = P_2^T \hat{S}_b P_2$ and $\bar{S}_w = P_2^T \hat{S}_w P_2$. It is easy to verify that \bar{S}_b is semi-positive definite and \bar{S}_w is positive definite in R^s . Thus we can calculate v_{r+1}, \dots, v_{r+s} following the Fisher LDA steps. Similarly, we can obtain the optimal regular discriminant vectors $W_r = [w_{r+1}, \dots, w_{r+s}]$ by using the mapping (4). From the irregular discriminant vectors W_{ir} and the regular discriminant vectors W_r , we can obtain the discriminant feature vector z of the sample x :

$$z = W^T x, \text{ where } W = [W_{ir}, W_r].$$

Based on the above descriptions in case 2, the CLDA algorithm can be described as follows:

Algorithm 1. The complete discriminant analysis

- step 1: work out the orthogonal eigenvectors p_1, \dots, p_l of the total class scatter matrix S_t corresponding to positive eigenvalues.
- step 2: Let $P = (p_1, \dots, p_l)$ and $\hat{S}_w = P^T S_w P$, $\hat{S}_b = P^T S_b P$, work out the orthogonal eigenvectors $\phi_1, \dots, \phi_r, \phi_{r+1}, \dots, \phi_{r+s}$ of \hat{S}_w .
- step 3: Let $P_1 = (\phi_1, \dots, \phi_r)$ and $\tilde{S}_b = P_1^T \hat{S}_b P_1$, work out the orthogonal eigenvectors v_1, \dots, v_r of \tilde{S}_b , calculate the irregular discriminant vectors W_{ir} by Eq.(2).
- step 4: Let $P_2 = (\phi_{r+1}, \dots, \phi_{r+s})$ and $\bar{S}_b = P_2^T \hat{S}_b P_2$, $\bar{S}_w = P_2^T \hat{S}_w P_2$, work out the optimal discriminant vectors v_{r+1}, \dots, v_{r+s} by the Fisher LDA, calculate the regular discriminant vectors W_r by Eq.(4).

step 5: Let $W = [W_{ir}, W_r]$ as projection to form the feature extractor.

III. THE FUZZY COMPLETE LINEAR DISCRIMINANT ANALYSIS

Inspired by [22-24], in this subsection, we may employ some statistic to incorporate the distribution information. Having this in mind, we look at the fundament results available in the setting of fuzzy nearest neighbor classifiers.

Given a sample set $X = \{x_1, x_2, \dots, x_n\}$, a fuzzy c -class partition of these vectors specify the membership degrees of each sample corresponding to each class. Thus the membership matrix $U = [u_{ij}]$, ($i = 1, 2, \dots, c$, $j = 1, 2, \dots, n$) can be obtain by FKNN (fuzzy k -nearest-neighbor) introduced in [18] and satisfies $\sum_{i=1}^c u_{ij} = 1$, $0 < \sum_{j=1}^n u_{ij} < n$. A detail computation example of the membership degrees can be found in [17]. In this paper, we calculate the membership degree matrix with Euclidean, seclidean, cityblock, Chebychev, cosine, correlation and Hamming distance. The computations of the membership degrees can be realized through a sequence of steps:

- Step 1: Compute the distance matrix between pairs of feature vectors in the training.
- Step 2: Set diagonal elements of this matrix to infinity (practically place large numeric values there).
- Step 3: Sort the distance matrix (treat each of its column separately) in an ascending order. Collect the 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).
- Step 4: Compute the membership grade to class i for j -th pattern using the expression proposed in [18].

$$u_{ij} = \begin{cases} 0.51 + 0.49(n_{ij}/k), & i = \text{the label of} \\ & \text{the } j \text{ th pattern,} \\ 0.49(n_{ij}/k), & \text{otherwise.} \end{cases}$$

In the above expression, n_{ij} stands for the number of the neighbors of the j -th data (pattern) that belong to the i -th class. Meanwhile in step 1, we adopt the suitable distance metric to calculate distance matrix by experiment. Taking into account the fuzzy membership degree, the mean vector of each class is calculated as follows:

$$\tilde{m}_i = \frac{\sum_{j=1}^n u_{ij} x_j}{\sum_{j=1}^n u_{ij}}, \quad i = 1, 2, \dots, c. \quad (6)$$

Incorporating the membership degrees, the between-class, within-class and the total class fuzzy scatter matrix of samples can be redefined as follows:

$$\begin{aligned} S_{bF} &= \sum_{i=1}^c \sum_{j=1}^n u_{ij} (\tilde{m}_i - m) (\tilde{m}_i - m)^T, \\ S_{wF} &= \sum_{i=1}^c \sum_{j \in N_i} u_{ij} (x_j - \tilde{m}_i) (x_j - \tilde{m}_i)^T, \\ S_{tF} &= S_{bF} + S_{wF}, \end{aligned} \quad (7)$$

where $m = \frac{1}{n} \sum_{i=1}^n x_i$ is the total mean of all samples. Then we will extract discriminant vectors based on the fuzzy scatter matrices. From Eqs.(6), (7), we can find that the fuzzy scatter matrices take into account the distribution information of samples by incorporating the membership degree. But in many small sample size problems, the fuzzy scatter matrix S_{wF} or S_{tF} may be still singular. Thus we adopt the same strategy in CLDA. First, we apply PCA to lower the dimension from m to l . Second, in the PCA subspace, we calculate the membership degree matrix and redefine the fuzzy scatter matrix using the membership degree matrix following Eq.(7). Last we calculate the irregular and regular discriminant in the null space and range space of S_{wF} . Therefore we call the new technique as Fuzzy-CLDA. Based on the above descriptions, the Fuzzy-CLDA algorithm can be described as follows:

Algorithm 2. The fuzzy complete discriminant analysis

- step 1: work out the orthogonal eigenvectors p_1, \dots, p_l of the total class scatter matrix S_t corresponding to positive eigenvalues.
- step 2: PCA transformation is implemented on original space, let $P = (p_1, \dots, p_l)$ and $\check{S}_t = P^T S_t P$.
- step 3: calculate the membership degree matrix U by the FKNN algorithm and the total class mean m in the PCA subspace.
- step 4: According to m and U work out the between-class, within-class and total class fuzzy scatter matrices in PCA subspace.
- step 5: Following the steps 2-5 of **Algorithm 1** to calculate the irregular and regular discriminant vectors.
- step 6: (Recognition): Project all samples into the obtained optimal discriminant vectors and classify.

IV. EXPERIMENTS

In this section, in order to demonstrate the effectiveness of Fuzzy-CLDA, we compare Fuzzy-CLDA with CLDA, UWLDA, FLDA, Fuzzy Fisherface, FIFDA on 3 different data sets from the UCI data sources. The characteristics of the three data sets can be found from (<http://archive.ics.uci.edu/ml/datasets>). Data 1 is a subset of from the Libras Movement Data Set, which contains 150 examples with 90 dimensions and 3 classes. Data 2 is a subset of the Optical Recognition of Handwritten Digits Data Set, which contains 150 examples with 48 dimensions and 3 classes. Data 3 is a subset of the Synthetic Control Chart Time Series Date Set, which contains 150 examples with 60 dimensions and 3 classes. All data sets are randomly split to the train set and test set with the ratio 1:4. Experiments are repeated 25 times to obtain mean prediction error rate as a performance measure, the nearest centroid classifier (NCC) is adopted to classify the test samples by using L2 norm. Our experiment is conducted in twofold. First, we discuss the effectiveness of the various distance metrics in the membership degree matrix; second, we compare the Fuzzy-CLDA with other approaches. All the experiments are performed on a Pentium 2.52GH with 2G

RAM and programmed in the MATLAB language (version 7.0).

A. The effectiveness of the distance metric in computing the membership degree matrix

In order to demonstrate the effectiveness of the various distance metrics in computing the membership degree, we calculate the distance matrix by the k -nearest neighbor algorithm with Euclidean, seculdean, cityblock, Chebychev, cosine, correlation and Hamming distance; while the optimal value k is estimated by cross-validation. The main result can be found in table 1. According to Table 1, we have the following conclusion: (1) In the data 1, the misclassification error rate is least when the correlation and cosine distance metrics are used in the k -nearest neighbor algorithm, moreover the optimal value k is 3. In the data 2, the misclassification error rate is least when the seculdean and Hamming distance metrics are used in the k -nearest neighbor algorithm, moreover the optimal value k is 2, 3, 4. In the data 3, the misclassification error rate is least when the cityblock distance is used in the k -nearest neighbor algorithm, moreover the optimal value k is 3. (2) For Fuzzy-CLDA, the distance metric in computing the distance matrix is very important, and the optimal distance metric is deferent for different data set. (3) For the k -nearest neighbor algorithm in FKNN, the optimal k is different in one distance metric.

B. Comparison with other approaches

To compare Fuzzy-CLDA with CLDA, UWLDA, FLDA, Fuzzy Fisherface and FIFDA, we choose the best distance metric and optimal value k from table 1. Moreover we choose the optimal value q as the power of the weighted function in UWLDA and FLDA through cross-validation. In the data 1, the optimal value q in UWLDA and FLDA is -2 and -3, respectively. In the data 2, the optimal value q in UWLDA and FLDA is -3 and -3, respectively. In the data 3, the optimal value q in UWLDA and FLDA is -4 and -3, respectively. We illustrate the effectiveness of Fuzzy-CLDA with misclassification rate and the reduced dimensionality, the main result can be found in table 2. According to Table 2, we have the following conclusion: (1) The reduced dimensionality by Fuzzy-CLDA is much lower than the original data. (2) In the date 1 and data 3, the misclassification error rate of the reduced dimensionality data preprocessed by Fuzzy-CLDA is much lower than the original data and other feature extraction methods; in the data 2, the misclassification error rate of the reduced dimensionality data preprocessed by Fuzzy-CLDA is not lower than the original data, however it is much lower than other methods. (3) The distance metric option in computing the membership degree matrix is better than the Fuzzy Fisherface and FIFDA. So Fuzzy-CLDA is outperforms all the other methods to some extent.

V. CONCLUSION

In this paper, we present Fuzzy-CLDA by combining CLDA and the membership degrees of samples. We redefine the fuzzy

scatter matrices by incorporating the membership degrees which relates the samples distribution information. Based on the the fuzzy scatter matrices, we calculate the irregular and regular discriminant vectors to reduce the dimensionality of the original data. In contrast to LDA and its extensions, Fuzzy-CLDA takes into account the samples distribution information and can be directly apply to the small sample size problems, thus it is more general for discriminant analysis to some extent. From table 1, we can see the distance metrics in computing the distance matrix and the optimal k in FKNN are important, which infect the membership degrees of samples and the classification results. From table 2, we can see Fuzzy-CLDA is outperform in the reduced dimensions, meanwhile the reduced dimensionality data can keep better discriminant information from the error rate. Furthermore we will apply the feature extraction technique to the classical small sample size problems such as information retrieval, text classification, gene microarray data selection, identity-based biometric identification, Hyperspectral Image Processing, etc. Meanwhile we will study the kernel Fuzzy-CLDA with respect to various kernel functions.

REFERENCES

- [1] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1999.
- [2] I.T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, 2002.
- [3] S. Mika, "Kernel Fisher Discriminant," Ph.D. thesis, University of Technology, 2002.
- [4] R.A. Fisher, "The Use of Multip Measurements in Taxonomic Problems," *Annual of Eugenics*, vol.7, pp.179-188, 1936.
- [5] C.R. Rao, "The Utilization of Multiple Measurements in Problems of Biological Classification," *Journal of the Royal Statistical Society*, vol.10, pp.159-203, 1948.
- [6] Z. Jin, J.Y. Yang, Z.S. Hu and Z. Lou, "Face recognition Based on the Uncorrelated discriminant Transformation," *Pattern Recognition*, vol.34, pp.1405-1416, 2001.
- [7] Z. Jin, J.Y. Yang, Z.M. Tang and Z.S. Hu, "A theorem on the Uncorrelated Optimal Discriminant Vectors," *Pattern Recognition*, vol.34, pp.2041-2047, 2001.
- [8] S.W. Ji and J.P. Ye, "Generalized Linear Discriminant Analysis: A Unified Framework and Efficient Model Selection," *IEEE Transactions on Neural Networks*, vol.19, pp.1768-1782, 2008.
- [9] J.W. Lu, K. Plataniotis and A. Venetsanopoulos, "Regularization Studies of Linear Discriminant Analysis in Small Sample Size Scenarios with Application to Face Recognition," *Pattern Recognition Letters*, vol.26, pp.181-191, 2005.
- [10] T. Hastie, A. Buja and R. Trbshirani, "Penalized Discriminant Analysis," *The Annals of Statistics*, vol.23, pp.73-102, 1995.
- [11] P. Howland and H. Park, "Gneralizing Discriminant Analysis Using the Generalized Singular Value Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, pp.995-1006, 2004.
- [12] J.P. Ye and Q. Li, "A Two-Stage Linear Discriminant Analysis via QR-Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, pp.929-941, 2005.
- [13] J.P. Ye, "Computational and Theoretical Analysis of Null Space and Orthogonal Linear Discriminant Analysis," *Jouranal of Machine Learning Research*, vol.7, pp.1183-1204, 2006.
- [14] L.F. Chen, H.Y. Mark Liao, M.T. KO, J.C. Lin and G.J. Yu, "A New LDA-Based Face Recognition System Which can Solve the Small Sample Size Problem," *Pattern Recognition*, vol.33, pp.1713-1726, 2000.
- [15] H. Yu and J. Yang, "A Direct LDA Algorithm for High-dimensional Date—with Application to Face Recognition," *Pattern Recognition*, vol.34, pp.2067-2070, 2001.
- [16] J. Yang and J.Y. Yang, "Why can LDA Be Performed in PCA Transformed Space?," *Pattern Recognition*, vol.36, pp.563-566, 2003.

Table 1 The misclassification error rate with various distance metrics

error rate (%)	k	Euclidean	seuclidean	cityblock	Chebychev	cosine	correlation	Hamming
data 1	2	7.60	11.33	8.53	8.27	7.07	7.07	8.93
	3	7.60	11.20	7.73	9.20	7.60	7.07	9.20
	4	7.40	9.07	7.47	11.47	9.47	8.27	9.07
	5	7.48	9.33	7.33	12.27	9.07	9.33	8.93
	6	8.67	10.80	7.20	13.60	8.80	9.07	8.80
	7	7.73	13.47	7.73	14.00	9.60	9.20	8.80
	8	8.13	11.07	7.20	16.67	9.73	8.67	8.00
	9	9.20	11.60	7.73	16.27	11.20	12.00	8.00
	data 2	2	0.67	0.40	0.53	1.07	0.80	1.07
3		0.80	0.40	1.07	1.73	0.80	0.80	0.40
4		0.80	0.40	1.33	1.33	0.80	0.80	0.40
5		0.93	0.53	1.33	1.07	0.93	0.67	0.53
6		1.20	0.67	1.87	0.93	1.07	0.80	0.67
7		1.60	0.80	2.67	1.20	1.07	0.93	1.07
8		1.73	0.93	2.13	1.33	1.73	1.47	1.20
9		2.40	0.93	2.67	3.87	1.87	2.13	1.47
data 3		2	4.93	6.13	5.20	6.53	7.33	5.87
	3	5.47	6.13	4.67	7.07	9.60	8.40	6.13
	4	5.60	6.13	6.40	10.00	11.87	11.47	6.13
	5	6.53	6.13	6.00	11.87	14.53	13.20	6.13
	6	8.00	6.67	5.60	11.07	17.07	16.13	6.80
	7	8.00	6.67	5.87	10.53	19.47	19.87	7.47
	8	7.60	6.93	4.93	10.80	19.87	21.87	7.33
	9	8.27	6.67	5.73	12.13	21.73	21.47	6.93

Table 2 The misclassification error rate with various distance metrics

method	error rate(%)			dimension		
	data1	data2	data3	data1	data2	data3
Original	14.67	0.00	13.33	90	48	60
CLDA	37.87	2.53	11.33	79	35	49
UWLDA	8.40	1.47	6.53	4	7	4
FLDA	12.00	0.93	14.53	2	2	2
Fuzzy Fisherface	16.13	9.47	12.93	2	2	2
FIFDA	15.47	4.13	19.07	2	2	2
Fuzzy-CLDA	7.07	0.40	4.67	2	3	2

[17] J. Yang, F. Frangi, J.Y. Yang, D. Zhang and Z. Jin, "KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, pp.230-244, 2005.

[18] M. Loog, "Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.23, pp.762-766, 2001.

[19] R. Lotlikar and R. Kothari, "Fractional-Step Dimensionality Reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, pp.623-627, 2000.

[20] J.W. Lu, K.N. Plataniotis and A.N. Venetsanopoulos, "Face Recognition Using LDA-Based Algorithm," *IEEE Transactions on Neural Networks*, vol.14, pp.195-200, 2003.

[21] Y.X. Liang, C.R. Li, W.G. Gong and Y.J. Pan, "Uncorrelated Linear Discriminant Analysis Based on Weighted Pairwise Fisher Criterion," *Pattern Recognition*, vol.40, pp.3606-3615, 2007.

[22] K. Kwak and W. Pedrycz, "Face Recognition Using a Fuzzy Fisherface Classifier," *Pattern Recognition*, vol.38, pp.1717-1732, 2005.

[23] W.K. Yang, J.G. Wang, M.W. Ren, L. Zhang and J.Y. Yang, "Feature Extraction Using Fuzzy Inverse FDA," *Neurocomputing*, vol.72, pp.3384-3390, 2009.

[24] M. Ilbeygi, H. Shah-Hosseini, "Anovel Fuzzy Facial Expression Recognition System Based on Facial Feature Extraction from Face Images," *Engineering Applications of Artificial Intelligence*, vol.25, pp.130-146, 2012.