Local graph embedding based on maximum margin criterion via Fuzzy Set

Minghua Wan^{1,2,3,4}, Zhihui Lai², Guowei Yang¹, Zhangjing Yang^{1,2}, Fanlong Zhang^{1,2}, Hao Zheng⁴

1) (School of Technology, Nanjing Audit University, Nanjing 211815, China)

2) (College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China)

3) (Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education (Nanjing University of Science and Technology), Nanjing 210094, China)

4) (Key Laboratory of Trusted Cloud Computing and Big Data Analysis, Nanjing Xiaozhuang University, Nanjing, 211171, P.R. China)

Email:wmh36@sina.com

Abstract: Recently, many based on locally graph embedding algorithms are proposed for dimensional reduction in nonlinear data. However, these algorithms are not effective when dealing with face images affected by variations in illumination conditions, poses or perspectives and different facial expressions. So, distant data points are not deemphasized efficiently by locally graph embedding algorithms and it may degrade the performance of classification. In order to solve the aforementioned problem, this paper proposes a new efficient dimension reduction method—local graph embedding method based on maximum margin criterion via Fuzzy Set for face recognition. Firstly, the goal of this algorithm is preserved under nearest neighbor premise by constructing the fuzzy intrinsic graph and the fuzzy penalty graph. Secondly, two novel fuzzy Laplacian scatter matrix are calculated using Fuzzy K-Nearest Neighbor (FKNN) in the proposed method. Finally, Maximum Margin Criterion (MMC) is used to avoid the “small size sample” problem. The results of face recognition experiments on the ORL, YALE and AR face databases demonstrate the effectiveness of our proposed method.

Key words: Locally linear embedding (LLE); Dimension reduction; Face recognition; Maximum margin criterion; Local graph embedding;

1 INTRODUCTION

Many locality preserving-based algorithms have been active areas of research because of their potential applications in computer vision, machine learning and pattern recognition fields.

Among them, methods include locally linear embedding (LLE) [1, 2], isometric feature mapping (ISOMAP) [3], local tangent space alignment (LTSA) [4] and Laplacian eigenmap (LE) [5, 6] are widely used. LLE is a linear subspace learning method, which constitutes local coordinates with the least constructed cost and maps them to a global one. Neighborhood preserving embedding (NPE) [7] optimally preserves the local neighborhood reconstruction relationships to find a low-dimensional embedding on the original data manifold. Locality preserving projections (LPP) [8, 9] can find an embedding space that preserves local information, which derived from Laplacian Eigenmap. Recently, Yan et al. [10] proposed linear graph embedding framework (LGE) to unite above manifold learning algorithms. Wan et al. [11] proposed LGE/MMC to solve small size sample problem in LGE algorithms.

However, LGE and LGE/MMC methods can not solve different facial expressions and variations in illumination conditions of face images in the real world. At the same time, distant points are not deemphasized efficiently by LGE and LGE/MMC and it may degrade the performance of classification. How can we completely 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. The fuzzy neighbor membership degree can efficiently handle the vagueness and ambiguity of samples being degraded by poor illumination, facial expression variations and distant points. In other words, the fuzzy neighbor membership degree helps to pull the near neighbor samples in same class nearer and nearer and repel the far neighbor samples of different classes farther and farther. So, the novel fuzzy neighbor graphs based on the fuzzy neighbor membership degree can better characterize the compactness and separability.

Many studies have been carried out for fuzzy image filtering, fuzzy image segmentation and fuzzy edge detection by taking advantage of the technology of fuzzy sets [12] in many problems of pattern recognition and image processing [13-16]. To solve these problems, in this paper we investigate local graph embedding method based on maximum margin criterion [17] and the fuzzy set theory for dimensional reduction.

The novelties of our method come from the following perspectives:

1. The proposed method constructs two new fuzzy graphs using FKNN , where it is important to maintain the original neighbor relations for neighboring data points of the same class and also

crucial to keep away neighboring data points of different classes. So, the novel fuzzy neighbor graphs based on the fuzzy neighbor membership degree can better characterize the compactness and separability.

2. Through the fuzzy neighbor graphs, the proposed algorithm has lower sensitivities to the sample variations caused by varying illumination, facial expression variations and distant points distant points. So, the class of a new test point can be more reliably predicted by the nearest neighbor criterion, owing to the locally discriminating nature.

3. The proposed method avoids the small sample size problem since it does not need to compute any matrix inversion; therefore, much computational time would be saved for feature extraction. However, the proposed method needs a few CPU time more than LPP and LLE. It is because that the proposed method builds two fuzzy graphs.

The rest of this paper is organized as follows: We review the ideas of several methods in section 2. In Section 3, we propose the idea of the proposed algorithm in detail. In section 4, we introduce the connections between LLE, NPE and the proposed algorithm. Experiments are presented to demonstrate the effectiveness of the proposed algorithm on face recognition in section 5. Finally, we give concluding remarks and a discussion of future work in Section 6.

2 OUTLINES OF FKNN, LPP AND MMC

In this section, we briefly review some basic subspace algorithm. Let us consider linear transformation mapping the set of N sample $X = \{x_1, x_2, \dots, x_N\}$, $x_i \in R^D$ into a d -dimensional feature space $Y = \{y_1, y_2, \dots, y_N\}$, where $y_i \in R^d$ and $D > d$. The new feature vectors $y_i \in R^d$ are defined by the following linear transformation:

$$y_i = V^T x_i, \quad i = 1, \dots, N, \quad (1)$$

where $V \in R^{D \times d}$ is a transformation matrix

2.1 Fuzzy K-Nearest Neighbor (FKNN)

In FKNN algorithm, Fuzzy K-nearest neighbor class produces the corresponding degrees of class membership. The computations of the membership degree can be realized through three

steps:

Step1: Set diagonal elements of this Euclidean distance matrix to infinity by computing the Euclidean distance matrix between pairs of feature vectors in training set.

Step2: 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 concerned with ‘ k ’ neighbors, this returns a list of ‘ k ’ integers).

Step3: Compute the membership degree to class ‘ i ’ for j^{th} pattern using the expression proposed in the literature [21].

$$u_{ij} = \begin{cases} 0.51 - 0.49 n_{ij}/k & \text{if } i = \text{the same as} \\ & \text{label of the pattern} \\ 0.49 \times (n_{ij}/k) & \text{if } i \neq \text{the same as the } j^{\text{th}} \\ & \text{label of the pattern} \end{cases} \quad (1)$$

For instance, if there are n_{ij} neighbors of the pattern that belong to the same category, the membership grade is kept close to 0.51. Otherwise the membership grade is close to 0.49.

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. As usual, u_{ij} satisfies two obvious properties:

$$\sum_{i=1}^c u_{ij} = 1, \quad 0 < \sum_{j=1}^N u_{ij} < N \quad (2)$$

where c is the number of class and N is the number of training samples in class i . Therefore, the fuzzy membership matrix U can be achieved with the result of FKNN.

$$U = [u_{ij}], \quad i = 1, 2, \dots, c, \quad j = 1, 2, \dots, N \quad (3)$$

2.2 Linear preserving projection (LPP)

The objective function of LPP[8-9] is defined as:

$$\min \sum_{i,j} \|y_i - y_j\| W_{ij} \quad (4)$$

where the similarity matrix W of LPP defined as:

$$W_{ij} = \begin{cases} 1, & \|x_i - x_j\|^2 < \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

110 W can be Gaussian weight of Euclidean distance using k -neighborhood, and $\|\bullet\|$ means the L_2
 111 norm.

112 After some matrix analysis steps, the minimization problem becomes

$$\begin{aligned} 113 \quad & \arg \min_V V^T XLX^T V \\ 114 \quad & \text{s.t. } V^T XDX^T V = 1 \end{aligned} \quad (6)$$

115 where $X = [X_1, X_2, \dots, X_N]$ is the training space of size $n \times N$, and D is a diagonal matrix
 116 whose entries are column or row sums of S . $L = D - W$ is the Laplacian matrix.

117 The optimal d projection vectors that minimizes the objective function can be computed by the
 118 minimum eigenvalues solutions to the generalized eigenvalues problem

$$119 \quad XLX^T V_i = \lambda_i XDX^T V_i \quad (7)$$

120 **2.3 Maximum margin criterion (MMC)**

121 The MMC [16] is based on the difference of between-class scatter matrix and within-class scatter
 122 matrix, which is defined as follows:

$$123 \quad J_s(w) = \text{tr}(V^T (S_b - \alpha S_w) V) \quad (8)$$

124 where the parameter α is a nonnegative constant which balances the relative merits of
 125 maximizing the between-class scatter to the minimization of the within-class scatter. The
 126 between-class scatter matrix S_b and within-class scatter matrix S_w can be denoted as

$$127 \quad S_b = \frac{1}{n} \sum_{i=1}^C n_i (m_i - m_0)(m_i - m_0)^T \quad (9)$$

$$128 \quad S_w = \frac{1}{n} \sum_{i=1}^C \sum_{j=1}^{n_i} (\mathbf{x}_j^i - m_i)(\mathbf{x}_j^i - m_i)^T \quad (10)$$

129 where n_i is the number of training samples in class i . In class i , the j^{th} training sample is denoted
 130 by $\mathbf{x}_j^i$, the mean vectors of training samples in class i is denoted by $m_i = (1/n_i) \sum_{k=1}^{n_i} \mathbf{x}_k^i$ and the mean
 131 vector of all training samples is $m_0 = (1/n) \sum_{k=1}^n \mathbf{x}_k$.

132

3 Local graph embedding based on maximum margin criterion via Fuzzy Set

3.1 The idea of the proposed method

The proposed method constructs two new fuzzy graphs using FKNN, where it is important to maintain the original neighbor relations for neighboring data points of the same class and also crucial to keep away neighboring data points of different classes. To avoid singular, motivated by the idea of MMC, the proposed method seeks to minimize the difference, rather than the ratio, between the fuzzy locality preserving between-class scatter and fuzzy locality preserving within-class scatter.

3.2 Fuzzy locality preserving within-class scatters

To begin with, we propose to minimize the fuzzy local scatter compactness of each data point by linear coefficients that reconstruct the data point from other points. The algorithm can be described in four steps.

1. K_c -nearest neighbors of each data points x_i using Euclidean distances.
2. Calculating the reconstructing weight matrix $W = [w_{ij}]_{N \times N}$ by minimizing the reconstruction error, which reconstructs each point x_i from its K_c .

$$\min J_L(W) = \sum_{i=1}^N \left\| x_i - \sum_{j=1}^{K_c} w_{ij} x_j \right\|^2 \quad (11)$$

where $w_{ij} = 0$ if x_i and x_j are not neighbors, and the rows of W sum to 1: $\sum_{j=1}^{K_c} w_{ij} = 1$.

After repeating the first step and the second step are performed on all the N data points, we can calculate the reconstruction weights to construct a weight matrix $W = [w_{ij}]_{N \times N}$.

3. *Compute affinity fuzzy weights.* Specify the affinity matrix W_{fuzzy}^G

$$W_{ij}^G = \begin{cases} u_{ij}^G W, & i \in \{x_{i1}, x_{i2}, \dots, x_{ik_c}\} \\ 0, & else \end{cases} \quad (12)$$

where, $u_{ij}^G = \begin{cases} 0.51 + 0.49 \times (n_{ij} / K_c) \\ 0.49 \times (n_{ij} / K_c) \end{cases}$.

4. Reconstructing represented y_i to obtain weight matrix W^G using the reconstruction error function:

$$\min J_L(Y) = \sum_{i=1}^N \left\| y_i - \sum_{j=1}^N w_{ij}^G y_j \right\|^2 \quad (13)$$

where y_i is the output of x_i , y_j is a neighbor of y_i .

Considering the map in Eq. (1), the objective function reduces to

$$\begin{aligned} J_L(V) &= \sum_{i=1}^N \left\| y_i - \sum_{j=1}^N w_{ij}^G y_j \right\|^2 = \sum_{i=1}^N \text{tr} \left\{ \left(y_i - \sum_{j=1}^N w_{ij}^G y_j \right) \left(y_i - \sum_{j=1}^N w_{ij}^G y_j \right)^T \right\} \\ &= \text{tr} \left\{ \sum_{i=1}^N \left(y_i - \sum_{j=1}^N w_{ij}^G y_j \right) \left(y_i - \sum_{j=1}^N w_{ij}^G y_j \right)^T \right\} \\ &= \text{tr} \left\{ Y \left(I - (W^G)^T \right) \left(I - (W^G)^T \right)^T Y^T \right\} \\ &= \text{tr} \left\{ Y \left(I - W^G \right)^T \left(I - W^G \right) Y^T \right\} \\ &= \text{tr} \left\{ V^T X M X^T V \right\} \end{aligned} \quad (14)$$

where $M = \left(I - W^G \right)^T \left(I - W^G \right)$.

3.3 Fuzzy locality preserving between-class scatters

To begin with, we propose to maximize the sum of pairwise squared distances between outputs if they have different labels. So, maximize fuzzy locality preserving between-class scatter of samples can be described in four steps.

1. Constructing the similarity matrix W_{ij}^p as follows:

$$W_{ij}^p = \begin{cases} 1, & \text{if } x_i \text{ is in the } K_p \text{ nearest from different classes of } x_j \\ 0, & \text{otherwise} \end{cases}$$

2. Compute affinity fuzzy weights. Specify the affinity matrix $W_{fuzzy}^{G'}$.

$$W_{ij}^{G'} = \begin{cases} u_{ij}^G W_{ij}^p, & \text{if } x_i \text{ is in the } K_p \text{ nearest from different classes of } x_j \\ 0, & \text{else} \end{cases}$$

$$\text{where } u_{ij}^{G'} = \begin{cases} 0.51 + 0.49 \times (n_{ij} / K_p) \\ 0.49 \times (n_{ij} / K_p) \end{cases}$$

3. Complete the embedding

$$\max J_G(Y) = \sum_{i=1}^N \sum_{j=1}^N \|y_i - y_j\|^2 \quad (15)$$

Considering the map Eq.(1), the objective function reduces to

$$\begin{aligned} J_G(V) &= \sum_i \sum_j \|y_i - y_j\|^2 W_{ij}^{G'} \\ &= \sum_i \sum_j \|V^T x_i - V^T x_j\|^2 W_{ij}^{G'} \end{aligned} \quad (16)$$

$$= 2V^T X(D^p - W^p)X^T V$$

$$= 2V^T XL^p X^T V$$

3.4 Criterion of the proposed method

At last, when the fuzzy locality preserving between-class scatter and the fuzzy locality preserving within-class scatter have been constructed, an intuitive motivation is to find a common projection that minimizes between-class scatter $J_L(V)$ and maximizes within-class scatter $J_G(V)$ at the same time. Actually, we can obtain such a projection by the following multi-object optimized problem, that is:

$$\begin{cases} \min tr \{V^T X M X^T V\} \\ \max tr \{V^T X L^p X^T V\} \end{cases} \quad (17)$$

$$\text{s.t. } V^T X X^T V = I$$

At last, it can be changed into the following constrained problem:

$$\min tr \{V^T X (M - \mu L^p) X^T V\} \quad (18)$$

$$\text{s.t. } V^T X X^T V = I$$

where μ is an adjustable parameter to balance between-class scatter and within-class scatter.

Eq. (18) can be solved by Lagrangian multiplier method:

$$L(V, \lambda) = \{V^T X (M - \mu L^p) X^T V - \lambda_i (V^T X X^T V - I)\} = 0 \quad (19)$$

where λ_i is the Lagrangian multiplier. Thus we get:

$$X (M - \mu L^p) X^T V = \lambda_i X X^T V \quad (20)$$

where V_i is generalized eigenvector correspondingly to generalized eigenvalue λ_i .

3.5 Comparisons of computation and memory complexities

In Table 1, we compare the computational and the memory space complexities of the six methods. Here m and n is the number of the rows and the columns of the image matrix. d , K_c , K_p , M and N is the number of the projection vectors, the testing and the training samples, respectively.

Table 1 Computation and memory complexities

Method	Complexity		
	Time (training)	Time (testing)	Memory
PCA	$O(mnN^2)$	$O(dmnN + dmnM)$	$O(mnN)$
LDA			
MMC			
LPP	$O(mnN^2 + mnN K_c^3)$	$O(dmnN + dmnM)$	$O(mnN)$
LLE	$O(mnN^2 + mnN K_c^3)$		
The proposed method	$O(mnN^2 + mnN K_c^3 + mnN K_p^3)$		

In Table 1, for PCA, LDA, MMC, LPP, LLE and the proposed method, since we need to perform $O(mnN+mnM)$ tests when using the nearest neighbor rule for classification and for each test it has the time complexity of $O(d)$, the testing time is $O(dmnN+dmnM)$. The memory cost is determined by the size of the matrices of the associated eigen equations, which is $O(mnN)$. The training time complexity depends on both the size of the matrices in the eigen equations and the number of the projection vectors that are required to be computed. For PCA, LDA and MMC, this is $O(mnN^2)$. For the LPP and LLE method, an extra time cost to construct the within-class similarity matrix, i.e., $O(mnNK_c^3)$, will be taken into account. However, the proposed method is constructed the within-class similarity matrices and between-class similarity matrices, which is $O(mnNK_c^3 + mnNK_p^3)$.

4 Connection between the proposed method and NPE

In this Section, the proposed method seems to be formally similar to NPE from form. However, the proposed method is also obviously different from it. We discuss to investigate the similarity

and the difference between the proposed method and NPE.

In NPE, the minimization problem reduces to finding V :

$$\min_{V^T XX^T V = I} \text{tr}\{V^T XMX^T V\} \quad (21)$$

where the matrix XMX^T is symmetric and semi-positive definite.

The transformation matrix V that minimizes the objective function is given by the minimum eigenvalue solution to the following generalized eigenvector problem:

$$XMX^T V_i = \lambda_i XX^T V_i \quad (22)$$

In the proposed method, it is formulated as the following constrained minimization problem:

$$\min_{U^T XX^T U = I} \text{tr}\{V^T X(M - \mu L^p)X^T V\} \quad (23)$$

Thus we have:

$$X\hat{M}X^T V_i = \lambda_i XX^T V_i \quad (24)$$

where $\hat{M} = M - \tilde{M}$, $\tilde{M} = \mu L^p$. It is easy to see that NPE is a special case of the proposed method (i.e. when $\mu = 0$).

5 Experiments and results

To evaluate the effectiveness of the proposed method algorithm, we systematically compare it with existing methods such as PCA, LDA, MMC, LPP and LLE algorithm in different databases: ORL, YALE and AR. In our experiments, we varied intraclass nearest neighbors $K_C = l - 1$ (l is training images from each class), where within-class samples are well clustered in the observation space [9]. We also test the impact of the nearest-neighbor parameter K_p on the recognition performance. If the value of K_p is too small, it is very difficult to preserve the topologic structure in the low-dimensional feature space. On the contrary, if the value of K_p is too large, it is very difficult to depict the assumption of local linearity in the high-dimensional feature space. We change the value of K_p (K_p from 2 to 20 with step 2) on the ORL, Yale face databases and the value of K_p (K_p from 5 to 50 with step 5) on the AR database, respectively. Euclidean distance and nearest neighborhood classifier are used in all the experiments. The

experiments were carried out on the same PC (CPU: P4 2.8 GHz, RAM: 1024 MB).

5.1 Experiments on the ORL Database

The ORL face database [18] contains images from 40 individuals, each providing 10 different images where the pose, face expression and sample size vary. The facial expressions and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in this scale of up to about 10 percent. All images normalized to a resolution of 56×46 . Fig. 1 shows sample images of one person from ORL face database.

Fig.1 Images of one person on the ORL database

We test the recognition performances of the six methods: PCA, LDA, MMC, LPP, LLE and the proposed method. In this experiments, we randomly selected l images (l varies from 2 to 6) to form the training sample set from the image gallery of each individual. The remaining $10-l$ images are used for testing. We independently run 50 times for each l images. In the PCA phase of LDA, MMC, LPP, LLE and the proposed method, we keep 95 percent image energy. Table 2 shows the average recognition accuracy (%) of different algorithms on the ORL face database and the corresponding standard deviations and dimensions.

Table 2 The average recognition accuracy (%) of different algorithms on the ORL face database and the corresponding standard deviations and dimensions

l Methods	2	3	4	5	6
PCA	72.25 ± 2.61 (22)	81.62 ± 0.65 (22)	83.98 ± 1.00 (22)	85.41 ± 1.30 (22)	85.41 ± 1.98 (22)
LDA	76.35 ± 1.07 (28)	83.40 ± 1.69 (28)	84.13 ± 2.03 (28)	86.07 ± 1.16 (28)	88.60 ± 0.78 (28)
MMC	75.01 ± 1.77 (26)	83.96 ± 0.48 (28)	85.37 ± 2.48 (28)	87.47 ± 0.85 (28)	90.49 ± 1.17 (26)
LPP	73.52 ± 0.05 (42)	80.28 ± 0.78 (44)	89.18 ± 1.42 (44)	91.13 ± 0.82 (40)	92.12 ± 1.21 (38)
LLE	69.78 ± 0.82 (32)	72.66 ± 0.73 (30)	76.32 ± 1.14 (36)	78.85 ± 1.15 (20)	88.44 ± 1.55 (28)
the proposed method	79.30 ± 0.83 (46)	88.61 ± 1.79 (42)	93.53 ± 1.82 (50)	95.84 ± 1.23 (48)	97.36 ± 1.24 (50)

Fig.2. shows the average recognition rates (%) of the proposed method versus the corresponding varied dimension when the 4 images per classes were randomly selected for training on the ORL face databases.

Fig.2. The average recognition rates (%) of the proposed method versus the corresponding varied dimension when the 4 images per classes were randomly selected for training on the ORL face databases.

5.2 Experiments on the YALE Database

The YALE face database [19] contains 165 gray scale images of 15 individuals, each individual has 11 images. The images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink). Fig. 3 shows sample images of one person from YALE face database.

Fig.3 Images of one person on the YALE database

In this experiment, each image was manually cropped and resized to 50×40 in Yale database. In the PCA phase of LDA, MMC, LPP, LLE and the proposed method, we keep 95 percent image energy. In the experiments, we randomly selected l images (l varies from 2 to 6) to form the training sample set from the image gallery of each individual. The remaining $11-l$ images are used for testing. We independently run 50 times for each l . Table 3 shows the average recognition

accuracy (%) of different algorithms on the YALE face database and the corresponding standard deviations and dimensions. Fig.4. shows the average recognition rates (%) of the proposed method versus the corresponding varied dimensions when the 6 images per classes were randomly selected for training on the YALE face databases.

Table 3 The average recognition accuracy (%) of different algorithms on the YALE face database and the corresponding standard deviations and dimensions

l Methods	2	3	4	5	6
PCA	77.29 ± 1.20 (18)	80.20 ± 1.27 (20)	83.99 ± 1.38 (18)	84.72 ± 1.24 (20)	86.21 ± 0.80 (22)
LDA	80.45 ± 1.48 (8)	84.55 ± 1.06 (8)	87.85 ± 0.45 (8)	87.20 ± 1.64 (8)	88.54 ± 0.82 (8)
MMC	80.58 ± 0.71 (12)	82.84 ± 0.88 (6)	85.87 ± 1.12 (6)	86.83 ± 0.37 (6)	87.79 ± 0.50 (6)
LPP	80.40 ± 1.80 (24)	85.79 ± 0.38 (22)	88.00 ± 0.50 (20)	89.57 ± 0.70 (18)	92.05 ± 1.23 (26)
LLE	83.55 ± 1.38 (14)	83.58 ± 0.59 (6)	85.90 ± 0.75 (6)	88.37 ± 1.63 (8)	88.97 ± 1.56 (8)
the proposed method	90.45 ± 0.75 (36)	91.48 ± 1.01 (36)	93.21 ± 1.24 (34)	95.34 ± 1.21 (32)	96.33 ± 1.21 (32)

Fig.4. The average recognition rates (%) of the proposed method versus the corresponding varied dimensions when the 6 images per classes were randomly selected for training on the YALE face databases.

5.3 Experiments on the AR Database

The AR face database [20] contains over 4,000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions, and occlusions. The pictures of 120 individuals (65 men and 55 women) were taken in two sessions (separated by two weeks) and each session contains 13 color images. The face portion of each image is manually cropped and then normalized to 50×40 pixels. These images vary in the following order: neutral expression, smiling, angry, screaming, left light on, right light on, all sides light on, wearing sun glasses, wearing sun glasses and left light on, wearing sun glasses and right light on. Fig. 5 shows sample images of one person from AR face database.

Fig.5 Images of one subject of the AR database. The first line and the second line images were taken in different time (separated by two weeks).

In this experiment, we randomly selected l images (l varies from 2 to 6) to form the training sample set from the image gallery of each individual. The remaining $20-l$ images are used for testing. We independently run 50 times for each l . In the PCA phase of LDA, MMC, LPP, LLE and the proposed method, the number of principle components is set as 150. The dimension steps are set to be 5 in final low-dimensional subspaces obtained by the six methods.

Table 4 The average recognition accuracy(%) of different algorithms on the AR face database and the corresponding standard deviations and dimensions

l	2	3	4	5	6
Methods					
PCA	66.68 ± 1.11 (85)	70.21 ± 1.62 (85)	77.60 ± 1.27 (85)	79.03 ± 0.81 (80)	81.54 ± 1.04 (85)
LDA	71.50 ± 0.71 (70)	75.58 ± 0.76 (70)	82.53 ± 0.81 (70)	87.12 ± 0.33 (70)	87.58 ± 0.75 (70)
MMC	69.39 ± 1.15 (80)	75.71 ± 0.59 (80)	82.98 ± 0.62 (85)	85.27 ± 0.94 (80)	87.86 ± 0.82 (80)
LPP	72.82 ± 0.81 (85)	78.78 ± 1.10 (85)	83.70 ± 0.65 (80)	88.75 ± 1.31 (80)	89.93 ± 0.48 (70)
LLE	70.40 ± 0.89 (85)	74.31 ± 1.16 (75)	83.35 ± 1.49 (70)	84.78 ± 0.76 (80)	86.98 ± 0.40 (85)
the proposed method	75.09 ± 1.49 (75)	82.42 ± 1.08 (70)	88.23 ± 0.89 (75)	90.17 ± 1.36 (75)	93.38 ± 1.20 (70)

Table 4 shows the average recognition accuracy(%) of different algorithms on the AR face database and the corresponding standard deviations and dimensions. Fig.6. shows the average recognition rates (%) of the proposed method versus the corresponding varied when the 4 images per classes were randomly selected for training on the AR face databases.

Fig.6. The average recognition rates (%) of the proposed method versus the corresponding varied when the 4 images per classes were randomly selected for training on the AR face databases.

5.4 Experiments on the AR Database

In addition, the average CPU time consumed for training, test and classification, and the maximal average recognition rates of the foregoing each method are given in Table 5-6. The proposed method achieves its maximal average recognition rate, and it needs less CPU time except the proposed method when compared to other methods on three databases, respectively.

Table 5 The average CPU time(s) consumed for training, test and classification, and the maximal average recognition rates (%) when the 2 images per person were randomly selected for training on the ORL face database

Methods	PCA	LDA	MMC	LPP	LLE	the proposed method
Recognition rate (%)	74.86	77.42	72.05	78.78	70.6	80.13
dim	(46)	(38)	(48)	(40)	(32)	(36)
CPU time(s)	0.281	0.360	0.316	0.350	0.365	0.415

Table 6 The average CPU time(s) consumed for training, test and classification, and the maximal average recognition rates (%) when the 2 images per person were used for training on the Yale face database

Methods	PCA	LDA	MMC	LPP	LLE	the proposed method
Recognition rate (%)	78.49	81.93	81.29	81.45	84.93	91.2
dim	(29)	(14)	(21)	(22)	(28)	(20)
CPU time(s)	0.155	0.151	0.140	0.152	0.163	0.196

5.4 Overall observations and discussions

Based on the experiments of three databases, we have several interesting observations:

[1] In Table 2-4, the proposed method gains much higher average recognition accuracy recognition rates than PCA, LDA, MMC, LPP and LLE on the three face databases, which shows that the proposed method can handle SSS problem very well. The reason may be that the proposed method can solve different facial expressions and variations in illumination conditions in three face databases by fuzzy set.

[2] In Fig. 2, 4 and 6, the average recognition rate (%) of the proposed method for a given number of dimensions is always higher than other five methods, which are demonstrated that the performance of the proposed method outperforms the other methods under the same condition. Different from other methods, it further shows that the proposed method can extract more discriminative features than the other methods.

[3] In Table 5-6, compared to other methods, the proposed method needs a few CPU time more than other methods for training, testing, and classification. It is because that 2DMED builds two graphs through fuzzy sets.

6 Conclusions

In this paper, we proposed a new supervised learning technique for feature extraction and recognition, namely local graph embedding method based on maximum margin criterion via fuzzy set. The fuzzy neighbor membership degree can efficiently handle the vagueness and ambiguity of samples being degraded by poor illumination, facial expression variations and distant points. The experimental results on ORL, YALE and AR face databases show that the proposed method

consistently outperforms the state-of-the-art techniques such as PCA, LDA, MMC, LLE and LPP.

ACKNOWLEDGEMENTS

This work is partially supported by the National Science Foundation of China under grant no. 61462064, 61203243, 61272077, 61202319, 61403188, 61563037, the China Postdoctoral Science Foundation under grant No. 2014T70453, 2013M530223 and Jiangsu Provincial Postdoctoral Science Foundation under grant No.1301095C, the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information (Nanjing University of Science and Technology), Ministry of Education (Grant No. 30920140122006), the University Natural Science Fund of JiangSu Province, China (Grant No.15KJB520018).

REFERENCES

- [1] S.T. Roweis, L.K. Saul (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500): 2323–2326.
- [2] L.K. Saul, S.T. Roweis (2003) Think globally, fit locally: unsupervised learning of low dimensional manifolds. *J. Mach. Learn. Res* 4: 119–155.
- [3] J.B. Tenenbaum, V. de Silva, J.C. (2000) Langford A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500): 2319–2323.
- [4] Z. Zhang, H. Zha (2004) Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal of Scientific Computing* 26(1): 313–338
- [5] M. Belkin and P. Niyogi, T. G. Dietterich, S. Becker, and Z. Ghahramani (2000) Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems* 14: 873-878
- [6] M. Belkin, P. Niyogi (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput* 15(6): 1373–1396.
- [7] X.F. He, D. Cai, S.C. Yan, H.J. Zhang (2005) Neighborhood preserving embedding. In: *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Beijing, china, October 2005, pp. 1208–1213
- [8] X. He and P. Niyogi (2003) Locality Preserving Projections. In: *Proceedings of the 17th Annual Conference on Neural Information Processing Systems*, Vancouver and Whistler, Canada, December 2003, pp. 153-160
- [9] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang (2005) Face Recognition Using Laplacianfaces. *IEEE Trans.*

Pattern Analysis and Machine Intelligence 27(3): 328-340

[10] S. Yan, D. Xu, B. Zhang, and H.-J. Zhang, Graph Embedding: A General Framework for Dimensionality Reduction, IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'05), vol. 2, pp. 830-837, 2005, San Diego.

[11] Wan, Minghua, Gai, Shan and Shao, Jie. "Local Graph Embedding Based on Maximum Margin Criterion (LGE/MMC) for Face Recognition.." Informatica (Slovenia) 36 , no. 1 (2012): 103-112.

[12] L.A. Zadeh, Fuzzy sets, Info. Control 8 (1965) 338–353.

[13] J.C. Bezdek, J. Keller, R. Krishnapuram, Fuzzy Models and Algorithms for *Pattern Recognition and Image Processing*, Kluwer Academic Publishers, Dordrecht, 1999.

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

[15] Minghua Wan, Ming Li, Zhihui Lai, Jun Yin, Zhong Jin. "Feature Extraction Based on Fuzzy Class Mean Embedding (FCME) with its Application to Face and Palm Biometrics" Machine Vision and Applications. Vol. 23, Issue 5 (2012) , Page 985–997

[16] Minghua Wan, Guowei yang, Zhong Jin. "Feature Extraction Based on Fuzzy Local Discriminant Embedding (FLDE) with Applications to Face Recognition" IET Computer Vision. Vol. 5, Issues 5, pp 301-308, 2011.

[17] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Transactions on Neural Networks 17 (1) (2006) 1157–1165.

[18] AT&T Laboratories Cambridge, The ORL database of faces, [_http://www.uk.research.att.com/facedatabase](http://www.uk.research.att.com/facedatabase).
Html

[19] Yale database, 1997. Available from: <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.

[20] A. Martinez, R. Benavente, The AR face database, CVC Tech. Rep. 24 (1998).