Machine Vision and Applications (2012) 23:985–997
DOI 10.1007/s00138-011-0365-5

ORIGINAL PAPER

Feature extraction based on fuzzy class mean embedding (FCME) with its application to face and palm biometrics

Minghua Wan · Ming Li · Zhihui Lai · Jun Yin · Zhong Jin

Received: 16 October 2010 / Revised: 7 July 2011 / Accepted: 22 July 2011 / Published online: 19 August 2011

Abstract In the local discriminant embedding (LDE) framework, the neighbor and class of data points were used to construct the graph embedding for classification problems. From a high-dimensional to a low-dimensional subspace, data points of the same class maintain their intrinsic neighbor relations, whereas neighboring data points of different classes no longer stick to one another. However, face images are always affected by variations in illumination conditions and different facial expressions in the real world. So, distant data points are not deemphasized efficiently by LDE and it may degrade the performance of classification. In order to solve above problems, in this paper, we investigate the fuzzy set theory and class mean of LDE, called fuzzy class mean embedding (FCME), using the fuzzy k-nearest neighbor (FKNN) and the class sample average to enhance its discriminant power in their mapping into a low dimensional space. In the proposed method, a membership degree matrix is firstly calculated using FKNN, then the membership degree and class mean are incorporated into the definition of the Laplacian scatter matrix. The optimal projections of FCME can be obtained by solving a generalized eigenfunction. Experimental results on the Wine dataset, ORL, Yale, AR, FERET face database and PolyU palmprint database show the effectiveness of the proposed method.

M. Wan (✉) · M. Li
School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China
e-mail: wmh36@sina.com

M. Wan · M. Li
Key Laboratory of Nondestructive Testing, Nanchang Hangkong University, Ministry of Education, Nanchang 330063, China

M. Wan · Z. Lai · J. Yin · Z. Jin
School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China

Keywords Local discriminant embedding (LDE) · Fuzzy k-nearest neighbor (FKNN) · Intrinsic neighbor relations · Graph embedding

List of symbols

c	The number of classes
x_i^j	The jth training sample in class i
y_i^j	The feature matrix of image matrix x_i^j
l_i^j	The x_i^j is labeled by some class label l_i^j
X	The set of the training samples
m	The total number of training samples
m_i	The total number of training samples in class i
$\bar{m}_i$	The class mean vector of training samples in class i
G_{fuzzy}	The fuzzy intraclass neighborhood graphs
G'_{fuzzy}	The fuzzy interclass neighborhood graphs
W_{fuzzy}^{G}	The fuzzy intraclass weight
$W_{\text{fuzzy}}^{G'}$	The fuzzy interclass weight
U_{ij}^{G}	The fuzzy intraclass membership matrix
$U_{ij}^{G'}$	The fuzzy interclass membership matrix
D^{G}	The fuzzy intraclass diagonal matrix
$D^{G'}$	The fuzzy interclass diagonal matrix
l	The number of training samples from each class
n	Sample dimension
n_{ij}	The number of the neighbors
d	The feature matrix dimension

1 Introduction

Feature extraction is an important research topic in computer vision, machine learning and pattern recognition fields. In the past several decades, many feature extraction methods have been proposed, in which principal component analysis

(PCA) [1] and linear discriminant analysis (LDA) [2] are two of the most fundamental feature extraction and dimensionality reduction methods. The optimal mapping of PCA is the leading eigenvectors of the data's total variance matrix associated with the leading eigenvalues. Thus, PCA preserved the total variance by maximizing the trace of feature variance, but PCA cannot preserve local information (local relationships within the data set) due to pursuing maximal variance. LDA is used to find the optimal set of projection vectors that maximize the determinant of the between-class scatter matrix and at the same time minimizing the determinant of the within-class scatter matrix [2]. In high-dimensional problems, like face recognition, the 2D face image matrices must be previously transformed into 1D image vectors column by column or row by row. However, concatenating 2D matrices into 1D vector often leads to a high-dimensional vector space, and the number of observations is small, usually tens or hundreds of samples. An intrinsic limitation of traditional LDA is that it fails to work when the within-class scatter matrix becomes singular. This is known as the small sample size (3S) problems, the undersampled or singularity problem. To overcome these weaknesses of PCA and LDA, other feature extraction approaches such as nullspace method [3], direct linear discriminant analysis (DLDA) [4], complete linear discriminant analysis (CLDA) [5], regularized linear discriminant analysis (RLDA) [6], independent component analysis (ICA) [7], Kernel principal component analysis (KPCA) [8], Kernel linear discriminant analysis (KLDA) [9] and Kernel local discriminant embedding (KLDE) [14] have been proposed.

However, PCA, LDA and their 2D versions fail to discover and preserve the local information on the manifold. A number of linear dimensionality reduction techniques have been developed to address this problem. Recently, He et al. [10] proposed a linear method named locality preserving projections (LPP) for dimensionality reduction that can preserve local relationships within the data set that lies on a lower dimensional manifold. Other nonlinear methods, such as isometric feature mapping (ISOMAP) [11], local linear embedding (LLE) [12], and Laplacian Eigenmap [13], have been proposed to find the intrinsic low-dimensional nonlinear data structures hidden in observation space. However, current manifold learning algorithms might be unsuitable for pattern recognition tasks in that they concentrate on representing the high-dimensional data with low-dimensional data instead of classification or that they only considered the locality and could not give a clear nonlinear map when applied to a new sample, such as ISOMAP and LLE. More recently, local discriminant embedding (LDE) [14] and marginal fisher analysis (MFA) [15] were proposed to overcome the drawbacks of LPP. LDE and MFA were developed by Chen et al., but the underlying ideas of which are almost the same: the neighbor and the class relations of data are utilized to construct the face space (subspace of the image space). Compared with LDA, MFA and LDE do not depend on the assumption that the data of each class is Gaussian distributed. Recently, the sparse representation-based classification (SRC) and its versions have been successfully used in face recognition [24–26]. However, these methods are very computationally expensive, and even prohibitive.

In the real world, face images are always affected by the variations in illumination conditions and different facial expressions. Fuzzy sets [17] can efficiently manage the vagueness and ambiguity of the face images being degraded by poor illumination component. By taking advantage of the technology of fuzzy sets, a number of studies have been carried out for fuzzy image filtering, fuzzy image segmentation, and fuzzy edge detection with an ultimate objective to cope with the factor of uncertainty being inherently present in many problems of image processing and pattern recognition [18,19,21]. Focusing on manifold learning and pattern classification, LDE achieves good discriminating performance by integrating the information of neighbor and class relations between data points. LDE incorporates the class information into the construction of embedding and derives the embedding for nearest-neighbor classification in a low-dimensional space, which learns the embedding for the submanifold of each class by solving an optimization problem. Nevertheless, distant points are not deemphasized efficiently by LDE and it may degrade the performance of classification.

To solve these problems, we investigate its extension and the fuzzy set theory, called fuzzy class mean embedding (FCME), using class mean of data points and the fuzzy k-nearest neighbor (FKNN) to enhance its discriminant power in their mapping into a low dimensional space. In the proposed method, a membership degree matrix is calculated using FKNN, then the membership degree is incorporated into the definition of the Laplacian scatter matrix. Significantly differing from the existing graph-based algorithms that two novel fuzzy neighbor graphs are constructed in FCME, 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 after the FCME. So, the class of a new test point can be more reliably predicted by the nearest neighbor criterion, owing to the locally discriminating nature. Through the fuzzy neighbor graphs, FCME algorithm has lower sensitivities to the sample variations caused by varying illumination, expression, viewing conditions and shapes.

The rest of the paper is structured as follows: in Sect. 2 we introduce LDE and FKNN. In Sect. 3, we propose the idea and describe FCME in detail. In Sect. 4, experiments on Wine database, ORL, Yale, AR, FERET face database and PolyU palmprint database are presented to demonstrate the effectiveness of FCME. Finally, we give concluding remarks.

2 Outline of LDE and FKNN

Let us consider a set of m sample $\{x_1, x_2, \ldots, x_m\}$ taking values in an n-dimensional image space, and assume that each image belongs to one of c classes. Let us also consider a linear transformation mapping the original n-dimensional space into an d-dimensional feature space, where $n > d$. The new feature vectors $y_k \in R^d$ are defined by the following linear transformation:

$$y_k = V^{\mathrm{T}} x_k, \quad k = 1, \ldots, m \tag{1}$$

where $V \in R^{n \times d}$ is a transformation matrix. The actual transformation in PCA includes a centering prior to the linear transform, where the data mean is subtracted.

The variables used in this paper are listed in list of symbols.

2.1 Local discriminant embedding (LDE)

Local discriminant embedding is a supervised subspace learning algorithm. In LDE class label l_i of $x_i (i = 1, \ldots, m)$ are used to determine a linear transformation matrix V such that:

$$y_i = V^{\mathrm{T}} x_i \tag{2}$$

The column vectors of $V = [v_1, v_2, \ldots, v_d]$ span a d-dimensional subspace. The aim of LDE is, in the low subspace, to keep neighboring points close if they have the same class label, whereas to prevent points of other classes from entering the neighborhood.

Graph G and G' are undirected graphs over data points, where G has edges between x_i and x_j if x_i and x_j belong to the same class, and G' has edges x_i and x_j if they belong to different classes, both in small neighborhoods. And W and W' are the weights in G and G, respectively. Its objective is to maximize the function:

$$J_{\mathrm{LDE}}(V) = \sum_{i,j} \| V^{\mathrm{T}} x_i - V^{\mathrm{T}} x_j \|^2 w'_{ij} \tag{3}$$

subject to

$$\sum_{i,j} \| V^{\mathrm{T}} x_i - V^{\mathrm{T}} x_j \|^2 w_{ij} = 1 \tag{4}$$

where

$$w'_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2/t), & \text{if } l_i \neq l_j \text{ and } i \in N^+_{K_C}(j) \\ & \quad \text{or } j \in N^+_{K_C}(i) \\ 0, & \text{else} \end{cases} \tag{5}$$

$$w_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2/t), & \text{if } l_i = l_j \text{ and } i \in N^+_{K_P}(j) \\ & \quad \text{or } j \in N^+_{K_P}(i) \\ 0, & \text{else} \end{cases} \tag{6}$$

where $N^+_{K_C}(i)$ indicates the index set of the K_C nearest neighbors of the sample x_i in the same class and $N^+_{K_P}(j)$ indicates the index set of the K_P nearest neighbors of the sample x_j in the different class. Of note, the affinity weights defined in (5) and (6) are derived from the heat kernel. Heat kernel is the fundamental solution to the heat equation on a particular domain with appropriate boundary conditions.

Equations (3) and (4) can be solved by Lagrangian multiplier method. The optimization can be reduced to the following generalized eigenvalue problem:

$$X(D' - W')X^{\mathrm{T}} v = \lambda X(D - W)X^{\mathrm{T}} v \tag{7}$$

where the elements of the matrix W' are w'_{ij}, the elements of the matrix W are w_{ij}. The elements of diagonal matrices D and D' are defined as $d_{ii} = \sum_j w_{ij}$ and $d'_{ii} = \sum_j w'_{ij}$, respectively.

2.2 Fuzzy K-nearest neighbor (FKNN)

In the real world, face images are always affected by variations in illumination conditions and different facial expressions. The fuzzy neighbor membership degree can efficiently handle the vagueness and ambiguity of samples being degraded by poor illumination, shape and facial expression variations. 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.

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.

With FKNN algorithm, the computations of the membership degree can be realized through a sequence of steps:

Step 1 Compute the Euclidean distance matrix between pairs of feature vectors in training set.

Step 2 Set diagonal elements of this 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 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 degree to class 'i' for jth pattern using the expression proposed in the literature [20]. 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. The nearer the neighbors, the greater the weights.

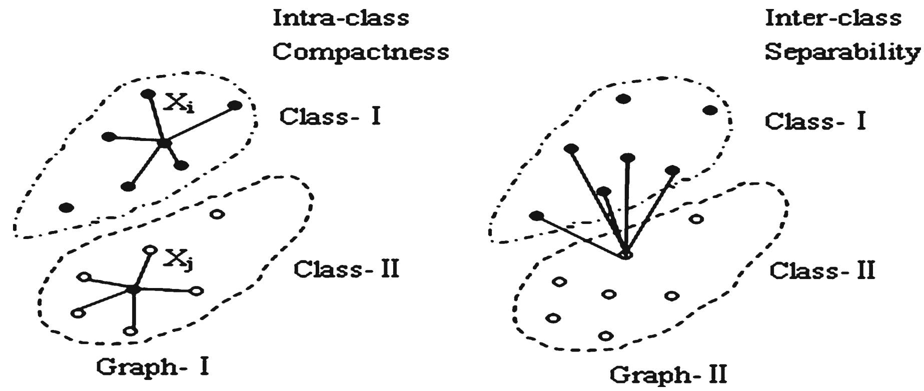


Fig. 1 The adjacency relationships of the intrinsic and penalty graphs for the local discriminant embedding (LDE) algorithm

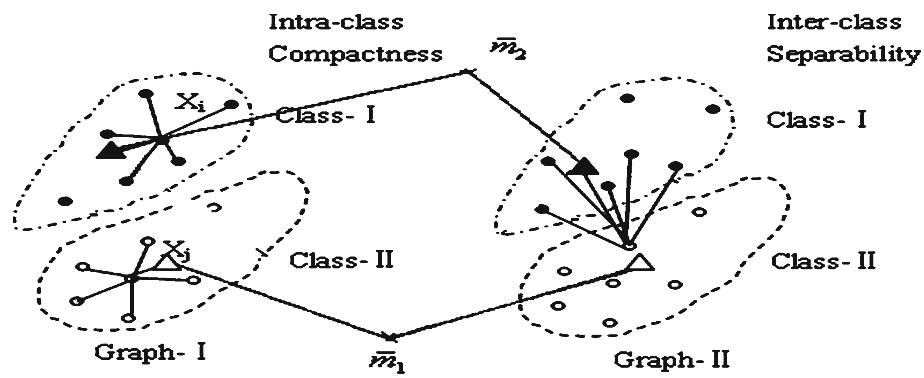


Fig. 2 The adjacency relationships of the intrinsic and penalty graphs for the fuzzy class mean embedding (FCME) algorithm

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

In the above expression, n_{ij} stands for the number of the neighbors of the jth data (pattern) that belong to the ith class. As usual, u_{ij} satisfies two obvious properties:

$$\sum_{i=1}^{c} u_{ij} = 1 \quad 0 < \sum_{j=1}^{N} u_{ij} < N \tag{9}$$

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

$$U = [u_{ij}], \quad i = 1, 2, \ldots c, \;\; j = 1, 2, \ldots N \tag{10}$$

3 The proposed FCME

Suppose there are c known pattern classes, $w_1, w_2, \ldots, w_c$, 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 jth training sample is denoted by x_i^j, the class mean vector of training samples in class i is denoted by $\bar{m}_i$, which called class mean vector. Taking into account the fuzzy membership degree, the mean vector of each class is:

$$\bar{m}_i = \frac{\sum_{j=1}^{m} u_{ij} x_i^j}{\sum_{j=1}^{m} u_{ij}} \tag{11}$$

For convenience of presentation, we first describe the steps of the FCME algorithm, and then justify them in detail. We have added c class mean vector $\bar{m}_i$, so there are $c + m$ data points $\{x_i\}_{i=1}^{m+c}$ are in $\Re^n$, and each x_i is labeled by some class label l_i. We also write the data matrix as $X = [x_1 x_2 \cdots x_m x_{m+1} \cdots x_{m+c}] \in R^n$. Figures 1 and 2 show the adjacency relationships of the intrinsic and penalty graphs, which respectively represent the LDE algorithm and the FCME algorithm. Studying the intrinsic and penalty graphs of LDE and FCME algorithm, we discover FCME graphs to be joined class mean vector $\bar{m}_i$ based on LDE graphs. In FCME graphs, each data points revolves around the class mean in same class, which causes more compact in the high dimension space and the data points tend the manifold distribution in the high dimensional space.

To solve these problems, we investigate its extension and the fuzzy set theory, called FCME, using class mean of data points and the FKNN to enhance its discriminant power in their mapping into a low dimensional space. Then, the proposed FCME can be realized by the following four steps.

Then, the proposed FCME can be realized by the following four steps.

1. *Construct fuzzy neighborhood graphs* Let G_{fuzzy} and G'_{fuzzy} denote two (undirected) graphs both over all data points. To construct G_{fuzzy},

$$G_{\text{fuzzy}}: \text{if} \left\{ \begin{array}{c} l_i = l_j \\ (i, j) \in m_i \\ i \in N^+_{K_C}(j) \text{ or } j \in N^+_{K_C}(i) \end{array} \right\} \tag{12}$$

For G'_{fuzzy},

$$G'_{\text{fuzzy}}: \text{if} \left\{ \begin{array}{c} l_i \neq l_j \\ (i, j) \notin m_i \\ i \in N^+_{K_P}(j) \text{ or } j \in N^+_{K_P}(i) \end{array} \right\} \tag{13}$$

where $N^+_{K_C}(i)$ indicates the index set of the K_C nearest neighbors of the sample x_i in the same class and $N^+_{K_P}(j)$ indicates the index set of the K_P nearest neighbors of the sample x_j in the different class.

2. Constructing the intraclass compactness and separability fuzzy G_{fuzzy}:

$$u^G_{ij} = \begin{cases} 0.51 + 0.49 \times (n_{ij}/K_C) \\ 0.49 \times (n_{ij}/K_C) \end{cases} \tag{14}$$

G'_{fuzzy} :

$$u^{G'}_{ij} = \begin{cases} 0.51 + 0.49 \times (n_{ij}/K_P) \\ 0.49 \times (n_{ij}/K_P) \end{cases} \tag{15}$$

3. *Compute affinity fuzzy weights* Specify the affinity matrix W^G_{fuzzy} of G_{fuzzy},

$$\begin{aligned} W^G_{ij} &= U^G_{ij} G_{\text{fuzzy}} \\ &= \begin{cases} u^G_{ij} \exp(-\|x_i - x_j\|^2/t), & i \in N^+_{K_C}(j) \\ 0, & \text{else} \end{cases} \end{aligned} \tag{16}$$

The other affinity matrix $W^{G'}_{\text{fuzzy}}$ of G'_{fuzzy} can be computed in the same way. In the proposed method, a membership degree matrix is calculated using FKNN, then the membership degree is incorporated into the definition of the Laplacian affinity matrix W and W' to get the fuzzy Laplacian affinity matrix W^G_{fuzzy} and $W^{G'}_{\text{fuzzy}}$, respectively. Significantly differing from the existing graph-based algorithms that two novel fuzzy affinity matrix are constructed in FCME, 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 than the Laplacian affinity matrix W and W' of LDE.

4. *Complete the embedding* Find the generalized eigenvectors $v_1, v_2, \ldots, v_d$ that correspond to the d largest eigenvalues in

$$X(D^{G'} - W^{G'})X^{\mathrm{T}}v = \lambda X(D^G - W^G)X^{\mathrm{T}}v \tag{17}$$

where D^G and $D^{G'}$ are diagonal matrices with diagonal elements

$$\begin{cases} d^G_{ii} = \sum_j w^G_{ij} \\ d^{G'}_{ii} = \sum_j w^{G'}_{ij} \end{cases} \tag{18}$$

The embedding of x_i is accomplished by

$$y_i = V^T x_i (i = 1, 2, \ldots, m). \tag{19}$$

where $V = [v_1, v_2, \ldots, v_d]$.

After the training by FCME, feature matrix of each image and a transformation matrix is obtained. Then a one-nearest neighbor classifier is used for classification.

Given two images x_1, x_2 represented by FCME feature vectors $y_1 = (y^1_1, y^2_1, \ldots, y^d_1)$ and $y_2 = (y^1_2, y^2_2, \ldots, y^d_2)$, then the dissimilarity $d(y_1, y_2)$ is defined as:

$$d(y_1, y_2) = \sum_{k=1}^{d} \|y^k_1 - y^k_2\| \tag{20}$$

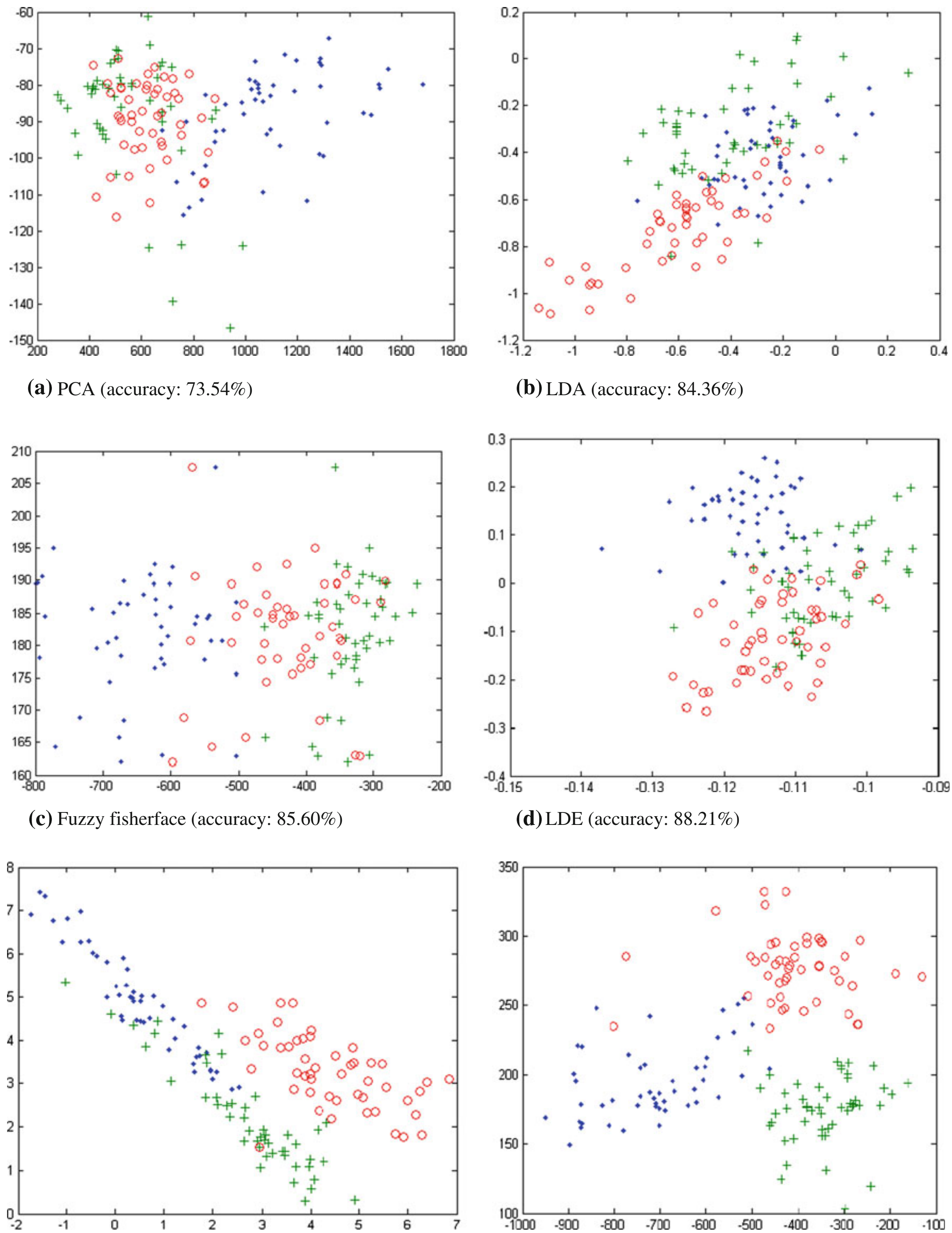
If the feature matrices of training images are $y_1, y_2, \ldots, y_m$, and each image is assigned to a class w_i. Then for a given test image y, if $d(y, y_l) = \min_j d(y, y_j)$ and $y_l \in w_i$, the resulting decision is $y \in w_i$.

4 Experiments and results

To evaluate the proposed FCME algorithm, we systematically compare it with the PCA, LDA, Fuzzy fisherface, LDE and KLDE algorithm on Wine database, ORL, Yale, AR, FERET face database and PolyU palmprint database. The Wine database was showed the effectiveness of FCME in constructing the two novel fuzzy graphs. The ORL database was used to evaluate the performance of FCME under conditions where the pose and sample size are varied. The Yale database was used to examine the system performance when both facial expressions and illumination are varied. The AR database was employed to test the performance of the system under conditions where there is a variation over time, facial expressions, and lighting conditions. The FERET face database was involved variations in facial expression, illumination and pose. The PolyU palmprint database was employed to test the performance of the system under conditions where there is a variation over time. 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 [23]. Euclidean distance and nearest neighborhood classifier are used in all the experiments.

4.1 Experiment on the WINE dataset from UCI: a toy example

Now we use Wine database, a real-life dataset from the UCI machine learning repository (http://archive.ics.uci.edu/ml),



(a) PCA (accuracy: 73.54%)

(b) LDA (accuracy: 84.36%)

(c) Fuzzy fisherface (accuracy: 85.60%)

(d) LDE (accuracy: 88.21%)

(e) KLDE (accuracy: 91.73%)

(f) FCME (accuracy: 95.85%)

Fig. 3 The points projected onto the 2D subspace learned by six methods and the corresponding recognition rate (shown in *parentheses*)

to show the effectiveness of FCME in constructing the two novel fuzzy graphs. Wine database consists of 178 samples of 3 classes. Every sample has 13 features. We select 48 samples per class in our experiments. Then first 8 out of 48 samples per class are selected for training. Here, we apply PCA, LDA, Fuzzy fisherface, LDE, KLDE and the proposed FCME for feature extraction. All the samples are projected onto the 2D subspace, which are shown in Fig. 3, respectively.

Fig. 4 Sample images of one person in the ORL face database

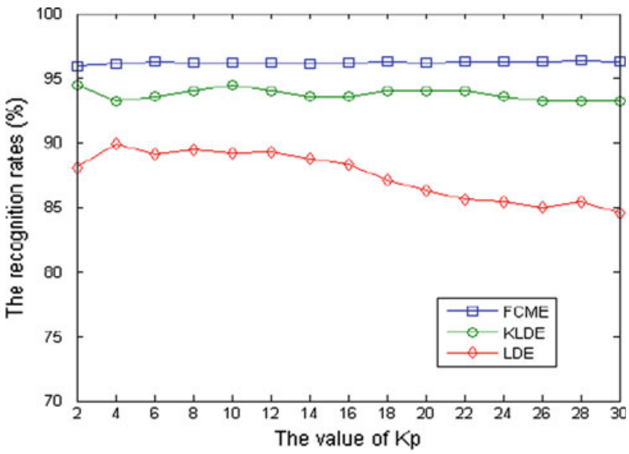


Fig. 5 The performances of LDE, KLDE and FCME are illustrated with the increase of K_P on the validation set of the ORL face database

According to the result shown in Fig. 3, the data points projected onto the 2D subspace learned by the proposed algorithm are clearly separated when compared with those of PCA, LDA, Fuzzy fisherface, LDE and KLDE. This indicates that FCME captures a more reasonable structure of the data.

4.2 Experiment on the face databases

The ORL database (http://www.uk.research.att.com/facedatabase.html) is used to evaluate the performance of FCME under conditions where the pose, face expression and sample size vary. The ORL face database contains images from 40 individuals, each providing 10 different images. 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°. Moreover, there is also some variation in the scale of up to about 10%. All images normalized to a resolution of 56×46. Figure 4 shows sample images of one person from ORL face database.

In our experiment, the first four samples of each class are used to compose the training set, the second three samples of each class compose the validation set, and the remaining three samples form the test set. In the PCA phase of LDA, Fuzzy fisherface, LDE, KLDE and FCME, we keep 90% image energy. In this experiment, we varied intraclass nearest neighbor parameter $K_C = 3$ and interclass nearest neighbor parameter K_P from 2 to 30 with an interval of 2, and the dimension of the extracted features vary from 2 to 50 with an interval of 2. The performances of LDE, KLDE and FCME are illustrated with the increase of K_P on the validation set in Fig. 5. It appears that FCME consistently outperforms LDE and KLDE. From these results, we choose the optimal parameter $K_P = 6$ for LDE, $K_P = 10$ for KLDE and $K_P = 4$ for FCME.

The Yale face database (http://www.cvc.yale.edu/projects/yalefaces/yalefaces.html) contains 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions. In our experiments, each image was manually cropped and resized to 100×80 pixels. Figure 6 shows sample images of one person. For computational effectiveness, we down sample it to 50×40 in this experiment. In the PCA phase of LDA, Fuzzy fisherface, LDE, KLDE and FCME, we keep 90% image energy.

In our experiment, the first three samples of each class are used to compose the training set, the second four samples of each class compose the validation set, and the remaining

Fig. 6 Sample images of one person in the Yale database

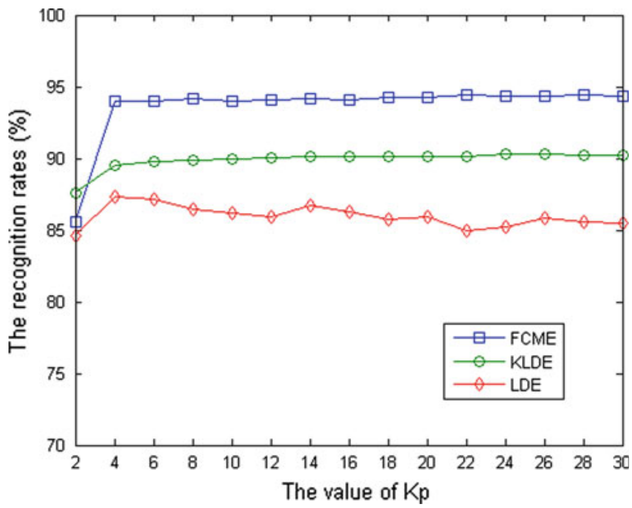


Fig. 7 The performances of LDE, KLDE and FCME are illustrated with the increase of K_P on the validation set of the Yale face database

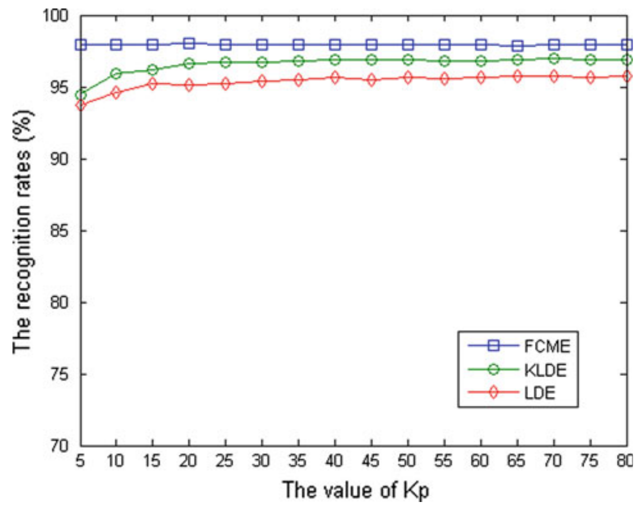


Fig. 9 The performances of LDE, KLDE and FCME are illustrated with the increase of K_P on the validation set of the AR face database

four samples form the test set. In the PCA phase of LDA, Fuzzy fisherface, LDE, KLDE and FCME, we keep 90% image energy. In this experiment, we varied intraclass nearest neighbor parameter $K_C = 2$ and interclass nearest neighbor parameter K_P from 2 to 30 with an interval of 2, and the dimension of the extracted features vary from 2 to 50 with an interval of 2. The performances of LDE, KLDE and FCME are illustrated with the increase of K_P on the validation set in Fig. 7. It appears FCME that consistently outperforms LDE and KLDE. From these results, we choose the optimal parameter $K_P = 4$ for LDE, $K_P = 16$ for KLDE and $K_P = 8$ for FCME.

The AR face database (http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html) 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 section contains 13 color images. The face portion of each image is manually cropped and then normalized to 50×40 pixels. The sample images of one person are shown in Fig. 8. These images vary as follows: (1) neutral expression, (2) smiling, (3) angry, (4) screaming, (5) left light on, (6) right light on, (7) all sides light on, (8) wearing sum glasses, (9) wearing sun glasses and left light on and (10) wearing sun glasses and right light on.

In our experiment, the first six samples of each class are used to compose the training set, the second seven samples of each class compose the validation set, and the remaining seven samples form the test set. In the PCA phase of LDA, Fuzzy fisherface, LDE, KLDE and FCME, we keep 90% image energy. The dimension steps are set to be five in final low-dimensional subspaces obtained by the six methods. In this experiment, we varied intraclass nearest neighbors parameter $K_C = 5$ and interclass nearest neighbors parameter K_P from 5 to 80 with an interval of 5, and the dimension of the extracted features vary from 5 to 150 with an interval of 5. Figure 9 shows the performances of LDE, KLDE and FCME are illustrated with the increase of K_P on

Fig. 8 Sample images of one subject of the AR database. *The first line and the second line images* were taken in different time (separated by 2 weeks)

Fig. 10 Samples of the cropped images from FERET database

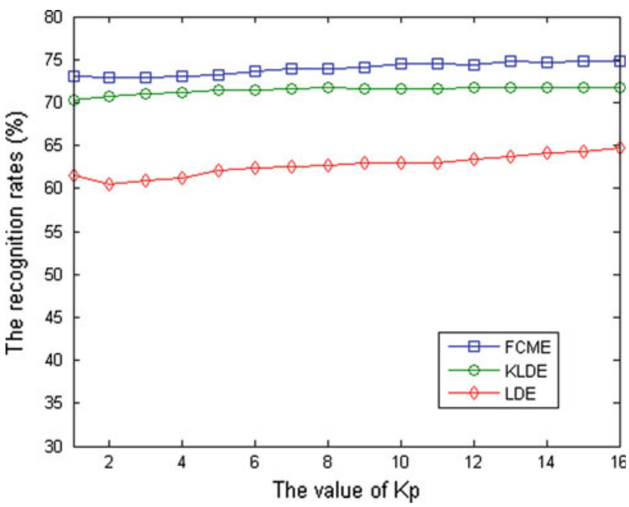


Fig. 11 The performances of LDE, KLDE and FCME are illustrated with the increase of K_P on the validation set of the FERET face database

Table 1 The recognition rates (%) of the three different methods on the test set of the ORL, Yale, AR and FERET face database

Databases	Method	LDE	KLDE	FCME
ORL	Recognition rates (%)	91. 65	93.20	96.15
	Dim	(24)	(26)	(28)
Yale	Recognition rates (%)	88.95	90.70	94.25
	Dim	(16)	(18)	(14)
AR	Recognition rates (%)	96.50	97.69	98.65
	Dim	(90)	(110)	(60)
FERET	Recognition rates (%)	65.36	72.80	75.18
	Dim	(50)	(50)	(45)

the validation set. It appears FCME that consistently outperforms LDE and KLDE. From these results, we choose the optimal parameter $K_P = 40$ for LDE, $K_P = 40$ for KLDE and $K_P = 20$ for FCME.

The FERET database (http://www.frvt.org/FERET/default.htm) includes 1,400 images of 200 distinct subjects, each subject has 7 images. The subset 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. Some facial portion images of one person are shown in Fig. 10.

In our experiment, the first three samples of each class are used to compose the training set, the second two samples of each class compose the validation set, and the remaining two samples form the test set. In the PCA phase of LDA, Fuzzy fisherface, LDE, KLDE and FCME, we keep 90% image energy. The dimension steps are set to be five in final low-dimensional subspaces obtained by the five methods. In this experiment, we varied intraclass nearest neighbors parameter $K_C = 2$ and interclass nearest neighbors parameter K_P from 5 to 80 with an interval of 5, and the dimension of the extracted features vary from 5 to 150 with an interval of 5. The performances of LDE, KLDE and FCME are illustrated with the increase of K_P on the validation set in Fig. 11. It appears FCME that consistently outperforms LDE and KLDE. From these results, we choose the optimal parameter $K_P = 40$ for LDE, $K_P = 30$ for KLDE and $K_P = 50$ for FCME.

In the first experiment, we obtain the recognition results of three methods on the test set based on these parameters, as listed in Table 1. Table 1 shows us that FCME performs better than LDE and KLDE.

In the second experiment, for further evaluating the performance of the proposed method, we randomly selected 4, 6, 5, 6 samples on the ORL, Yale, AR, FERET database from each class for training, while the remaining 6, 5, 15, 1 samples are used for testing, respectively. We run the system 50 times and obtain 50 different training and testing sample sets for performance evaluation on the ORL and Yale database. And we run the system ten times and obtain ten different training and testing sample sets for performance evaluation on the AR and FERET database. Based on the optimal parameters we obtain on the validation set in the foregoing experiment, we perform PCA, LDA, Fuzzy fisherface, LDE, KLDE and FCME. The maximal average recognition rates (%) and the corresponding dimensions (shown in parentheses) across 50 tests are listed in Table 2.

By comparing the recognition results in the columns of Table 2, we find that for all of the six feature extraction methods, the FCME achieves the better results than the other methods.

Figures 12, 13, 14 and 15 showed the variation of accuracy along different number of eigenvectors used and the recognition accuracy when the four, six, five, six samples per class are randomly selected for training set on the ORL, Yale, AR, FERET database. From four figures, we can see that FCME performs always better than the other three methods. The figures also demonstrate that the performance of the proposed method outperforms the other methods under the same condition, and it further shows that the proposed method can extract more discriminative features than the other methods.

Table 2 The maximal average recognition rates (%) of PCA, LDA, Fuzzy fisherface, LDE, KLDE and FCME and the corresponding dimensions (shown in parentheses) on the ORL, Yale, AR and FERET face database

Databases	Method	PCA	LDA	Fuzzy fisherface	LDE	KLDE	FCME
ORL	Recognition rates (%)	82.27	85.09	88.31	89.56	91.69	95.13
	Dim	(46)	(38)	(40)	(36)	(38)	(24)
Yale	Recognition rates (%)	87.01	89.36	92.35	93.95	95.26	97.92
	Dim	(46)	(14)	(16)	(20)	(20)	(20)
AR	Recognition rates (%)	79.84	87.45	87.62	91.86	92.65	96.71
	Dim	(150)	(115)	(125)	(85)	(80)	(75)
FERET	Recognition rates (%)	61.80	85.95	86.50	87.45	95.85	98.80
	Dim	(150)	(45)	(80)	(40)	(40)	(35)

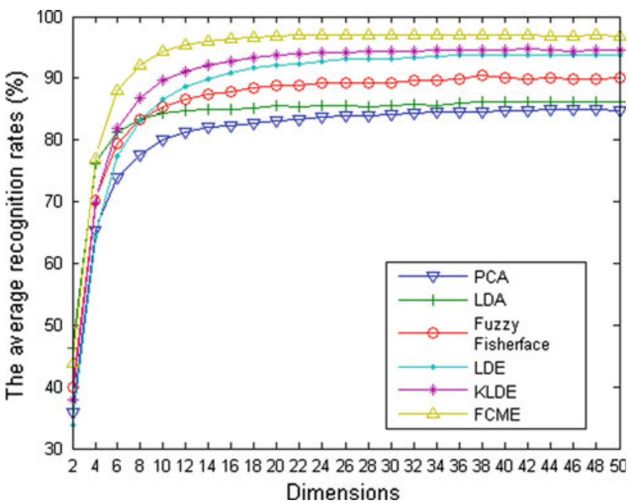


Fig. 12 The average recognition rates (%) of PCA, LDA, Fuzzy fisherface, LDE, KLDE and FCME versus the dimensions when the four images per class were randomly selected for training on the ORL face database

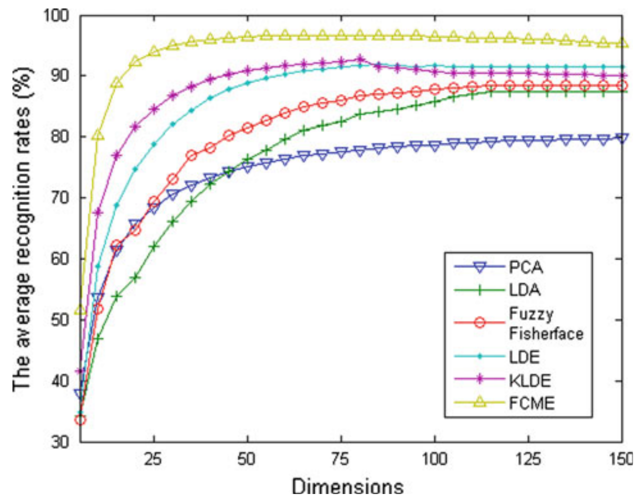


Fig. 14 The average recognition rates (%) of PCA, LDA, Fuzzy fisherface, LDE, KLDE and FCME versus the dimensions when the five images per class were randomly selected for training on the AR face database

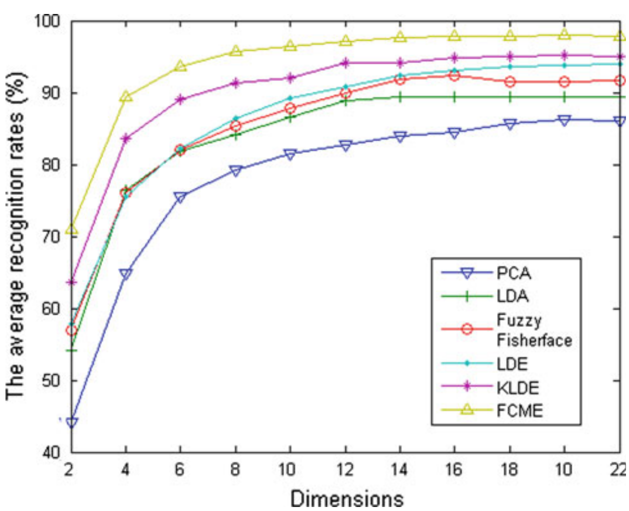


Fig. 13 The average recognition rates (%) of PCA, LDA, Fuzzy fisherface, LDE, KLDE and FCME versus the dimensions when the six images per class were randomly selected for training on the Yale face database

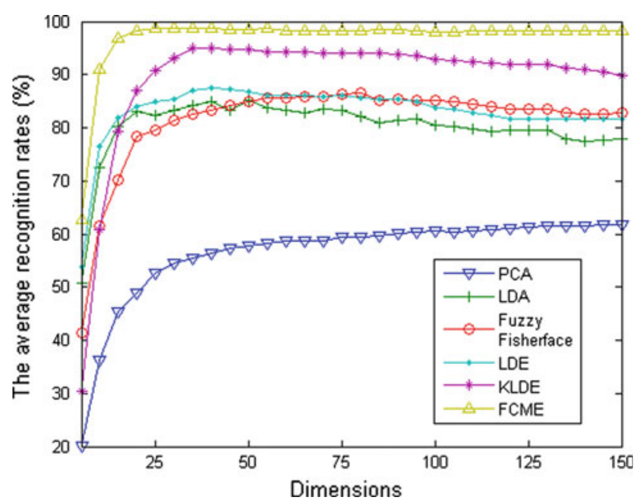


Fig. 15 The average recognition rates (%) of PCA, LDA, LDE and FCME versus the dimensions when the six images per class were randomly selected for training and the remaining one images per class for testing on the FERET face database

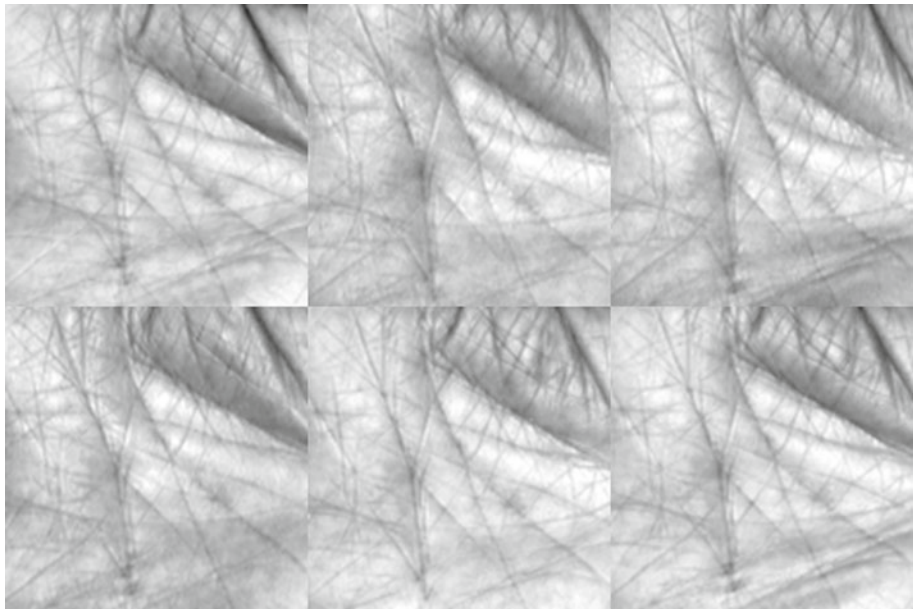


Fig. 16 Samples of the cropped images in the PolyU Palmprint database

Table 3 The maximal recognition rates (%) of PCA, LDA, Fuzzy fisherface, LDE, KLDE and FCME on the PolyU palmprint database and the corresponding dimensions when the first 300 samples are used for training and the remaining for test

Method	PCA	LDA	Fuzzy fisherface	LDE	KLDE	FCME
Recognition rates (%)	87.73	89.33	92.67	94.67	96.55	99.67
Dim	(125)	(140)	(90)	(95)	(90)	(45)

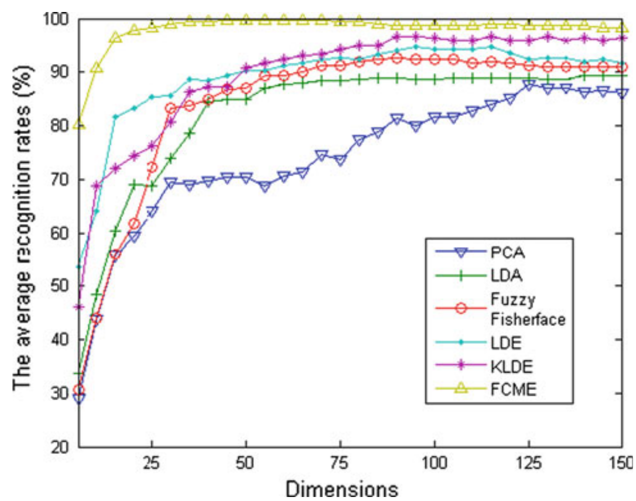


Fig. 17 The recognition rates (%) of PCA, LDA, Fuzzy fisherface, LDE, KLDE and FCME versus the dimensions when the first 300 images were used for training on the PolyU palmprint database. The dimension here is the number of eigenvectors

4.3 Experiment on the PolyU Palmprint database

The PolyU palmprint database contains 600 gray-scale images of 100 different palms with 6 samples for each palm (http://www4.comp.polyu.edu.hk/~biometrics/). Six samples from each of these palms were collected in two sessions, where the first three were captured in the first session and the other three in the second session. The average interval between the first and the second sessions is 2 months. In our experiments, the central part of each original image was automatically cropped using the algorithm mentioned in [22]. The cropped images were resized to 64 × 64 pixels and preprocessed using histogram equalization. Figure 16 shows some sample images of two palms. The maximal recognition rates of each method and the corresponding dimension are given in Table 3. As it is shown in Table 3, the top average recognition rate of FCME is significantly higher than the other methods.

According to the protocol of this database, the palmprint images are divided into 2 groups: 1 group is made up of 3 images of every palm from one session for a total of 300 images; the other group is made up of three images of every palm from the other session for a total of 300 images. Thus, for each palm class, there are three training samples and three test samples.

Seen from the Fig. 17, the FCME obtained the best recognition rates in a very low-dimension space. This experiment might imply that our method is more suitable for Palmprint recognition.

4.4 Overall observations and discussions

According to the experiments performed on the three face databases, the following conclusions can also be drawn:

- Differing from PCA, LDA and Fuzzy fisherface which attempt to preserve the global Euclidean structure, LDE, KLDE and FCME aim to discover the local geometric structure. Local structure based manifold learning algorithms are superior to the methods based on global structure.
- A common property of LDE, KLDE and FCME is that they both aim to discover the local geometric structure. But the recognition rates of FCME are significant higher than that of LDE and KLDE.
- FCME consistently outperforms PCA, LDA, Fuzzy fisherface, LDE and KLDE in spite of the variation of dimensions, which are shown in Tables 1 and 2.
- The average recognition rates (%) of FCME versus the dimensions is always higher than others five methods, which are shown in Figs. 12, 13, 14, 15 and 17.

5 Conclusion

In pattern recognition, feature extraction techniques are widely employed to reduce the dimensionality of data and to enhance the discriminatory information. In this paper, we develop a supervised discriminant technique, called FCME, using the FKNN to enhance its discriminant power in their mapping into a low dimensional space. Based on the class information, our approach achieves good accuracy by realigning the submanifolds and rectifying the neighbor relations in the embedding space. In FCME, two fuzzy graphs are constructed to characterize the within-class compactness and the between-class separability, 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 after the FCME. Experimental results show that FCME can capture a more reasonable structure of the data on the Wine database and outperforms other methods on the ORL, Yale, AR, FERET face database and PolyU palmprint database, respectively. The proposed method also discovers the local geometric structure which can reduce the sensitivity of the method to substantial variants between face images caused by large pose, expression or illumination variations.

Acknowledgments This work is partially supported by the National Science Foundation of China under grant no. 60632050, 90820306, 60873151, 60973098, 61005008 and 61005005, 60963002.

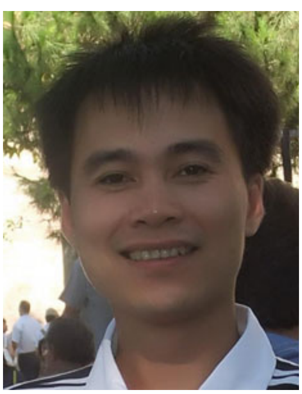
References

1. Turk, M., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–591 (1991)
2. Belhumeur, P.N., Hespanha, J.P., Kriengman, D.J.: Eigenfaces versus Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
3. Chen, L., Liao, H., Ko, M., Lin, J., Yu, G.: A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recognit. **33**(10), 1713–1726 (2000)
4. Yu, H., Yang, J.: A direct LDA algorithm for high-dimensional data-with application to face recognition. Pattern Recognit. **34**(11), 2067–2070 (2001)
5. Yang, J., Yang, J.-Y.: Why can LDA be performed in PCA transformed space? Pattern Recognit. **36**, 563–566 (2003)
6. Dai, D.Q., Yuen, P.C.: Regularized discriminant analysis and its application on face recognition. Pattern Recognit. **36**(1), 845–847 (2003)
7. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face recognition by independent component analysis. IEEE Trans. Neural Netw. **13**(11), 1450–1464 (2002)
8. Kim, K.I., Jung, K., Kim, H.J.: Face recognition using kernel principal component analysis. IEEE Signal Process. Lett. **9**(2), 40–42 (2002)
9. Zhang, Y., Gu, X., Liu, C.: Face recognition using kernel discriminant analysis. High Technol. Lett. **8**(12), 43–46 (2002)
10. He, X., Niyogi, P.: Locality Preserving Projections. In: Proceedings of the 16th Conference on Neural Information Processing Systems (2003)
11. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**, 2319–2323 (2000)
12. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**, 2323–2326 (2000)
13. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. **15**(6), 373–1396 (2003)
14. Chen, H., Chang, H., Liu, T.: Local discriminant embedding and itsvariants. CVPR (2005)
15. Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q., Lin, S.: Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. In: IEEE Transaction on Pattern Analysis and Machine Intelligence (T-PAMI), pp. 40–51 (2007)
16. Yang, W., Wang, J., Ren, M., Yang, J.: Feature extraction based on laplacian bidirectional maximum margin criterion. Pattern Recognit. **42**(11), 2327–2334 (2009)
17. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
18. Bezdek, J.C., Keller, J., Krishnapuram, R.: Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Kluwer Academic Publishers, Dordrecht (1999)
19. Kw, K.C., Pedry, W.: Face recognition using a fuzzy fisher classifier. Pattern Recognit. **38**(10), 1717–1732 (2005)
20. Keller, J.M., Gray, M.R., Givern, J.A.: A fuzzy *k*-nearest neighbour algorithm. IEEE Trans. Syst. Man Cybern. **15**(4), 580–585 (1985)
21. Yang, W., Wang, J., Ren, M., Zhang, L., Yang, J.: Feature extraction using fuzzy inverse FDA. Neurocomputing **72**(13–15), 3384–3390 (2009)
22. Zhang, D.: Palmprint Authentication. Kluwer, Dordrecht (2004)
23. Yang, J., Zhang, D., Yang, J., Niu, B.: Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics. IEEE Trans. Pattern Anal. Mach. Intell. **29**(4), 650–664 (2007)
24. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE PAMI **31**(2), 210–227 (2009)
25. Yang, M., Zhang, L.: Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In: ECCV (2010)
26. Yang, M., Zhang, L., Yang, J., Zhang, D.: Robust sparse coding for face recognition. CVPR (2011)

Author Biographies

Minghua Wan received his BS degree in automated institute from the Nanchang University of Aviation in 2003 and his MS and PhD degrees in pattern recognition and intelligent systems from the Nanjing University of Science and Technology (NUST) in 2007 and 2011, respectively. He is the author of more than 20 scientific papers in pattern recognition and computer vision. His current research interests include face recognition and detection, and image processing. E-mail: wmh36@sina.com.

Ming Li received his BS degree in electronic engineering from the Shanghai Jiaotong University, PRC in 1985 and his MS and PhD degrees in automatic control from the Nanjing University of Aeronautics and Astronautics, PRC in 1990 and 1997, respectively. He is a professor at the Nanchang Hangkong University. His current interests are in the areas of image processing and evolutionary computation.



Zhihui Lai received his BS degree in mathematics from the South China Normal University, MS degree from the Jinan University, China, in 2002 and 2005, respectively, and PhD degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology (NUST), Nanjing, China in 2011. His research interests include face recognition, image processing, pattern classification, neural networks, and human vision modelization and applications in the fields of intelligent robot research. E-mail: lai_zhi_hui@163.com.

Jun Yin received his BS degree in mathematics and PhD in pattern recognition and intelligence system from the Nanjing University of Science and Technology (NUST), Nanjing, China in 2006 and 2011, respectively. His current research interest includes pattern recognition, face recognition and machine learning.

Zhong Jin received his BS degree in mathematics, MS degree in applied mathematics and PhD degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology (NUST), China, in 1982, 1984 and 1999, respectively. He is a professor in the Department of Computer Science, NUST. He visited the Department of Computer Science and Engineering, the Chinese University of Hong Kong from January 2000 to June 2000 and from November 2000 to August 2001. He also visited the Laboratoire HEUDIASYC, UMR CNRS 6599, Universite de Technologie de Compiegne, France, from October 2001 to July 2002, as well as the Centre de Visio per Computador, Universitat Autonoma de Barcelona, Spain, as the Ramon y Cajal program Research Fellow. His current interests are in the areas of pattern recognition, computer vision, face recognition, facial expression analysis and content-based image retrieval. E-mail: zhong.jin@patternrecognition.cn.