Journal of Southeast University (English Edition) Vol. 24, No. 4, pp. 428 - 432 Dec. 2008 ISSN 1003—7985

Facial expression recognition based on fuzzy-LDA/CCA

Zhou Xiaoyan[1,2] Zheng Wenming[3] Zou Cairong[1,4] Zhao Li[1]

([1] School of Information Science and Engineering, Southeast University, Nanjing 210096, China)
([2] School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)
([3] Research Center for Learning Science, Southeast University, Nanjing 210096, China)
([4] Foshan University, Foshan 528000, China)

Abstract: A novel fuzzy linear discriminant analysis method by the canonical correlation analysis (fuzzy-LDA/CCA) is presented and applied to the facial expression recognition. The fuzzy method is used to evaluate the degree of the class membership to which each training sample belongs. CCA is then used to establish the relationship between each facial image and the corresponding class membership vector, and the class membership vector of a test image is estimated using this relationship. Moreover, the fuzzy-LDA/CCA method is also generalized to deal with nonlinear discriminant analysis problems via kernel method. The performance of the proposed method is demonstrated using real data.

Key words: fuzzy linear discriminant analysis; canonical correlation analysis; facial expression recognition

Linear discriminant analysis (LDA) [1] is one of the most popular feature extraction methods in the statistical pattern recognition field, which aims to find the optimal discriminant vectors that maximize the ratio of the between-class scatter to the within-class scatter of a given data set. LDA has been successfully used in many recognition problems such as face recognition, image retrieval, facial expression recognition, and so on. The traditional LDA method is always derived under the assumption that each training sample belongs to one class. However, there are some cases such as facial expression where each training sample may belong to more than one class category, so that the traditional discriminant analysis method may not be successfully used. For example, in the facial expression recognition task, each facial image may contain all the six basic facial expressions (happiness, sadness, surprise, anger, disgust and fear). Thus, it is not very reasonable to simply classify each facial image into only one of the six basic expressions. Recently, Kwak and Pedrycz[2] proposed a fuzzy fisherface approach for face recognition. They proposed to incorporate a gradual level of assignment as a membership grade to each class such that the discrimination helps to improve classification results. Compared with the traditional fisherface approach[3], however, the fuzzy fisherface approach, in nature, is still limited in the traditional LDA framework, and the formulations are also similar to those of the fisherface approach. The major difference between the fuzzy fisherface and the traditional fisherface lies only in the calculation of the mean of each class sample.

This paper proposes a novel fuzzy linear discriminant analysis method via canonical correlation analysis[4] (fuzzy-LDA/CCA) to recognize facial expressions. CCA is a method to correlate the linear relationships between two multidimensional variables. More specifically, let $\{\boldsymbol{x}, \boldsymbol{y}\} \in \mathbf{R}^{n \times m}$ be two random multidimensional variables, then the CCA method aims to find a pair of directions $\boldsymbol{\omega}_x$ and $\boldsymbol{\omega}_y$ such that the correlation $\rho(\boldsymbol{x}, \boldsymbol{y})$ between the two projections $\boldsymbol{\omega}_x^{\mathrm{T}} \boldsymbol{x}$ and $\boldsymbol{\omega}_y^{\mathrm{T}} \boldsymbol{y}$ is maximized. Discriminant analysis using CCA was proposed by Barker et al. [5]. However, their method only simply used a dummy matrix representing class membership, which may obtain the similar discriminant performance of LDA. Similar to the fuzzy fisherface method, the fuzzy K-nearest neighbor method (fuzzy K-NN) is used to design the class membership of each training sample. Therefore, the fuzzy-LDA/CCA method fully utilizes class membership information for the discriminant analysis. On the other hand, we use the least square regression (LSR) technique to establish the relationship between the input data and the corresponding class membership data, and then assign the class index according to the class membership data.

The kernel method has been popularly studied in recent years to solve the nonlinear feature extraction or classification problems[6]. The basic idea of the kernel-based learning algorithm is to map input data into a high dimensional reproducing Hilbert kernel space (RHKS) and then perform the same learning methods as those in the input data space, where the computational problems in RHKS can be solved via kernel method. The typical kernel based learning algorithms are kernel principal component analysis (KPCA) [7], generalized discriminant analysis (GDA) [8-9], and kernel canonical correlation analysis (KCCA) [4], etc. This paper also proposes the kernel version of the fuzzy-LDA/CCA method via the kernel method, for conducting the nonlinear discriminant analysis problems.

1 Canonical Correlation Analysis (CCA)

Let $\{\boldsymbol{x}, \boldsymbol{y}\} \in \mathbf{R}^{n \times m}$ denote a pair of random multidimensional variables. Then the goal of CCA is to find a pair of directions $\boldsymbol{\omega}_x$ and $\boldsymbol{\omega}_y$ such that the correlation $\rho(\boldsymbol{x}, \boldsymbol{y})$ between the two projections $\boldsymbol{\omega}_x^{\mathrm{T}}\boldsymbol{x}$ and $\boldsymbol{\omega}_y^{\mathrm{T}}\boldsymbol{y}$ is maximized[4], where

$$\rho(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\omega}_x, \boldsymbol{\omega}_y) = \frac{E\{\boldsymbol{\omega}_x^{\mathrm{T}}\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\omega}_y\}}{\sqrt{E\{\boldsymbol{\omega}_x^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}_x\} E\{\boldsymbol{\omega}_y^{\mathrm{T}}\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\omega}_y\}}} =$$

Received 2008-03-10.

Biographies: Zhou Xiaoyan (1980—), female, graduate; Zheng Wenming (corresponding author), male, doctor, professor, wenming_zheng@seu.edu.cn.

Foundation items: The National Natural Science Foundation of China (No. 60503023, 60872160), the Natural Science Foundation for Universities of Jiangsu Province (No. 08KJD520009), the Intramural Research Foundation of Nanjing University of Information Science and Technology (No. Y603).

Citation: Zhou Xiaoyan, Zheng Wenming, Zou Cairong, et al. Facial expression recognition based on fuzzy-LDA/CCA[J]. Journal of Southeast University (English Edition), 2008, 24(4): 428 - 432.

$$\frac{\boldsymbol{\omega}_x^{\mathrm{T}} E\{\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}\}\boldsymbol{\omega}_y}{\sqrt{\boldsymbol{\omega}_x^{\mathrm{T}} E\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}\boldsymbol{\omega}_x}\sqrt{\boldsymbol{\omega}_y^{\mathrm{T}} E\{\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\}\boldsymbol{\omega}_y}} \tag{1}$$

Let $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1,2,\dots,N}$ be N observations of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Then CCA aims to solve the following optimization problem:

$$\{\boldsymbol{\omega}_x^*, \boldsymbol{\omega}_y^*\} = \arg\max_{\boldsymbol{\omega}_x, \boldsymbol{\omega}_y} \rho(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\omega}_x, \boldsymbol{\omega}_y) = \arg\max_{\boldsymbol{\omega}_x, \boldsymbol{\omega}_y} \left\{ \frac{\boldsymbol{\omega}_x^{\mathrm{T}} \boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\omega}_y}{\sqrt{\boldsymbol{\omega}_x^{\mathrm{T}} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\omega}_x}\sqrt{\boldsymbol{\omega}_y^{\mathrm{T}} \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\omega}_y}} \right\} \tag{2}$$

where $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N\}$, $\boldsymbol{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_N\}$. The optimization of Eq. (2) can be solved using the Lagrangian approach, where the corresponding Lagrangian is

$$L(\boldsymbol{\omega}_x, \boldsymbol{\omega}_y, \lambda, \mu) = \boldsymbol{\omega}_x^{\mathrm{T}} \boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\omega}_y - \frac{\lambda(\boldsymbol{\omega}_x^{\mathrm{T}} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\omega}_x - 1)}{2} - \frac{\mu(\boldsymbol{\omega}_y^{\mathrm{T}} \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\omega}_y - 1)}{2} \tag{3}$$

Taking derivatives of L with respect to $\boldsymbol{\omega}_x$ and $\boldsymbol{\omega}_y$ and setting to zeros, we obtain

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\omega}_x} &= \boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\omega}_y - \lambda \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\omega}_x = \boldsymbol{0} \\ \frac{\partial L}{\partial \boldsymbol{\omega}_y} &= \boldsymbol{Y}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\omega}_x - \mu \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\omega}_y = \boldsymbol{0} \end{aligned} \tag{4}$$

From Eq. (4), we obtain

$$\mu = \lambda$$

$$\boldsymbol{\omega}_y = \frac{(\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}})^{-1} \boldsymbol{Y}\boldsymbol{X}^{\mathrm{T}}}{\mu} \boldsymbol{\omega}_x$$

and

$$\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}} (\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}})^{-1} \boldsymbol{Y}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\omega}_x = \lambda^2 \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\omega}_x \tag{5}$$

The generalized eigenequation (5) can be transformed into a symmetric eigenproblem by applying a complete Cholesky decomposition approach to matrix $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$[4].

2 Fuzzy-LDA/CCA

2.1 Formulations

Let $\{\boldsymbol{x}_i, s_i^j\}_{i=1,2,\dots,N; j=1,2,\dots,c}$ be a training data set with N samples belonging to c classes, where each training point $\boldsymbol{x}_i \in \mathbf{R}^n$ is associated with c coefficients $s_i^j (j = 1, 2, \dots, c)$ indicating the membership under the constraint $\sum_{j=1}^{c} s_i^j = 1$. Let $\boldsymbol{y}_i = \{s_i^1, s_i^2, \dots, s_i^c\}^{\mathrm{T}}$. Suppose that $\{(\boldsymbol{\omega}_x^i, \boldsymbol{\omega}_y^i)\}_{i=1}^{t}$ is the t pairs of directions of the two variables $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ using the CCA approach. Let

$$\boldsymbol{P}_x = \{\boldsymbol{\omega}_x^1, \boldsymbol{\omega}_x^2, \dots, \boldsymbol{\omega}_x^t\}, \quad \boldsymbol{P}_y = \{\boldsymbol{\omega}_y^1, \boldsymbol{\omega}_y^2, \dots, \boldsymbol{\omega}_y^t\} \tag{6}$$

Let $\boldsymbol{a}_i$ and $\boldsymbol{b}_i (i = 1, 2, \dots, N)$ be the projections of $\boldsymbol{x}_i$ and $\boldsymbol{y}_i (i = 1, 2, \dots, N)$ onto $\boldsymbol{P}_x$ and $\boldsymbol{P}_y$, respectively. Then we have $\boldsymbol{a}_i = \boldsymbol{P}_x^{\mathrm{T}} \boldsymbol{x}_i$, $\boldsymbol{b}_i = \boldsymbol{P}_y^{\mathrm{T}} \boldsymbol{y}_i$. Assume that there is a mapping function $f(\cdot)$ such that $f(\boldsymbol{a}) = \boldsymbol{b}$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are the projections of $\boldsymbol{x}$ and $\boldsymbol{y}$ onto $\boldsymbol{P}_x$ and $\boldsymbol{P}_y$, respectively. In this paper, we simply assume that f is a linear transformation since precisely solving the mapping function is a difficult task. Thus, we obtain that there exists a t by t matrix $\boldsymbol{P}$ such that

$$\boldsymbol{P}\boldsymbol{a} \approx \boldsymbol{b} \tag{7}$$

The least square regression approach is used to derive the expression of $\boldsymbol{P}$. Let

$$\varepsilon(\boldsymbol{P}) = \|\boldsymbol{P}\boldsymbol{a} - \boldsymbol{b}\|^2 = \boldsymbol{a}^{\mathrm{T}} \boldsymbol{P}^{\mathrm{T}} \boldsymbol{P}\boldsymbol{a} - \boldsymbol{a}^{\mathrm{T}} \boldsymbol{P}^{\mathrm{T}} \boldsymbol{b} - \boldsymbol{b}^{\mathrm{T}} \boldsymbol{P}\boldsymbol{a} + \boldsymbol{b}^{\mathrm{T}} \boldsymbol{b} \tag{8}$$

The optimization is performed by setting the partial derivative of $\varepsilon(\boldsymbol{P})$ with respect to $\boldsymbol{P}$ equal to zero:

$$\frac{\partial \varepsilon}{\partial \boldsymbol{P}} = 2\boldsymbol{P}\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}} - 2\boldsymbol{b}\boldsymbol{a}^{\mathrm{T}} = \boldsymbol{0} \tag{9}$$

From Eq. (9), we obtain that $\boldsymbol{P}\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}} = \boldsymbol{b}\boldsymbol{a}^{\mathrm{T}}$. Let $\boldsymbol{R}_{aa} = E\{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{a}_i \boldsymbol{a}_i^{\mathrm{T}}$ and $\boldsymbol{R}_{ba} = E\{\boldsymbol{b}\boldsymbol{a}^{\mathrm{T}}\} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{b}_i \boldsymbol{a}_i^{\mathrm{T}}$. Then we obtain that the transformation matrix $\boldsymbol{P}$ can be estimated by

$$\hat{\boldsymbol{P}} = \boldsymbol{R}_{ba} \boldsymbol{R}_{aa}^{-1} \tag{10}$$

Let $\boldsymbol{x}_{\text{test}}$ and $\boldsymbol{y}_{\text{test}}$ be a test sample and the corresponding class membership vector. Let $\boldsymbol{a}_{\text{test}}$ and $\boldsymbol{b}_{\text{test}}$ be the projections of $\boldsymbol{x}_{\text{test}}$ and $\boldsymbol{y}_{\text{test}}$ onto $\boldsymbol{P}_x$ and $\boldsymbol{P}_y$, respectively. Then we obtain

$$\hat{\boldsymbol{P}} \boldsymbol{a}_{\text{test}} = \boldsymbol{b}_{\text{test}}, \quad \boldsymbol{a}_{\text{test}} = \boldsymbol{P}_x^{\mathrm{T}} \boldsymbol{x}_{\text{test}}, \quad \boldsymbol{b}_{\text{test}} = \boldsymbol{P}_y^{\mathrm{T}} \boldsymbol{y}_{\text{test}} \tag{11}$$

From Eq. (11), we obtain $\hat{\boldsymbol{P}} \boldsymbol{P}_x^{\mathrm{T}} \boldsymbol{x}_{\text{test}} = \boldsymbol{P}_y^{\mathrm{T}} \boldsymbol{y}_{\text{test}}$. Thus, we obtain

$$\boldsymbol{y}_{\text{test}} = (\boldsymbol{P}_y \boldsymbol{P}_y^{\mathrm{T}})^{-1} \boldsymbol{P}_y \hat{\boldsymbol{P}} \boldsymbol{P}_x^{\mathrm{T}} \boldsymbol{x}_{\text{test}} \tag{12}$$

Let $\boldsymbol{y}_{\text{test}}^i$ denote the i-th element of $\boldsymbol{y}_{\text{test}}$, then the index of the most matched expression class of the test image is

$$\boldsymbol{c}^* = \arg\max_i \boldsymbol{y}_{\text{test}}^i \tag{13}$$

2.2 Assigning class membership using fuzzy *K*-nearest neighbor

The fuzzy-LDA/CCA method requires that each training point $\boldsymbol{x}_i$ is associated with a class membership vector. We adopt the fuzzy K-NN method[2, 10] to do this task. Considering that the fuzzy K-NN algorithm needs defining a distance metric to calculate the class membership of each data point, we use the following Euclidean norm as the distance metric $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ for any two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, where

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} (\boldsymbol{x}_i - \boldsymbol{x}_j)} = \sqrt{\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_i - 2\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j + \boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{x}_j} \tag{14}$$

Fuzzy K-NN algorithm

Input data: the training samples $\boldsymbol{x}_i$; the number of neighbors k.

Output data: the fuzzy class membership s_i^j.

1) Compute the distance, denoted by $d_{ij} = d(\boldsymbol{x}_i, \boldsymbol{x}_j)$, between any two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

2) Let $\boldsymbol{D}$ be an $N \times N$ matrix whose elements are composed of d_{ij}. Set the diagonal element of $\boldsymbol{D}$ to be infinite.

3) Sort the elements of each column of $\boldsymbol{D}$ in an ascending order. Collect the class labels of the patterns located in the closest neighborhood of the pattern under consideration(here we are concerned with “k” neighbors).

4) Compute the degree of membership of the j-th sample to the i-th class using the following approach: If i is equal to the class label of the j-th training sample, then $s_i^j = 0.51 + 0.49\,(n_i^j/k)$; otherwise, then $s_i^j = 0.49\,(n_i^j/k)$, where n_i^j stands for the number of the neighbors of the j-th sample belonging to the i-th class.

3 Fuzzy-KDA/KCCA Method

Let Φ be a nonlinear mapping that maps $\mathbf{R}^n$ from the input space into a Hilbert space F, i. e. , Φ: $\mathbf{R}^n \to F, \boldsymbol{x} \to \Phi(\boldsymbol{x})$, where the inner product of any two points $\Phi(\boldsymbol{x})$ and $\Phi(\boldsymbol{y})$ can be computed via the kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j): k(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\Phi(\boldsymbol{x}_i))^{\mathrm{T}}\Phi(\boldsymbol{x}_j)$. In this case, our goal turns to solving the pair directions of $\boldsymbol{\omega}_{\Phi(x)}$ and $\boldsymbol{\omega}_y$, such that the correlation $\rho(\Phi(\boldsymbol{x}), \boldsymbol{y})$ is maximized:

$$\rho(\Phi(\boldsymbol{x}), \boldsymbol{y}; \boldsymbol{\omega}_{\Phi(x)}, \boldsymbol{\omega}_y) = \frac{\boldsymbol{\omega}_{\Phi(x)}^{\mathrm{T}}\ \Phi(X)\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\omega}_y}{\sqrt{\boldsymbol{\omega}_{\Phi(x)}^{\mathrm{T}}\ \Phi(X)(\Phi(X))^{\mathrm{T}}\boldsymbol{\omega}_{\Phi(x)}}\ \sqrt{\boldsymbol{\omega}_y^{\mathrm{T}}\ \boldsymbol{YY}^{\mathrm{T}}\boldsymbol{\omega}_y}} \tag{15}$$

where $\Phi(X) = \{\Phi(\boldsymbol{x}_1), \Phi(\boldsymbol{x}_2), \ldots, \Phi(\boldsymbol{x}_N)\}$. The optimization problem in Eq. (15) can also be solved using the Lagrangian method. For more details, see Ref. [11]. Suppose that $\{(\boldsymbol{\omega}_{\Phi(x)}^i, \boldsymbol{\omega}_y^i)\}_{i=1}^t$ are the t pair directions of KCCA. Then according to Ref. [11], we know that $\boldsymbol{\omega}_{\Phi(x)}^i$ can be expressed as $\boldsymbol{\omega}_{\Phi(x)}^i = \Phi(\boldsymbol{x})\alpha_i$, where α_i is an N dimensional vector.

Now let $\boldsymbol{P}_{\Phi(x)} = [\boldsymbol{\omega}_{\Phi(x)}^1, \ldots, \boldsymbol{\omega}_{\Phi(x)}^t]$, $\boldsymbol{P}_y = [\boldsymbol{\omega}_y^1, \ldots, \boldsymbol{\omega}_y^t]$, and $\boldsymbol{A} = [\alpha_1, \ldots, \alpha_t]$. Then we obtain that $\boldsymbol{P}_{\Phi(x)} = \Phi(X)\boldsymbol{A}$. Let $\Phi(\boldsymbol{x}_{\text{test}})$ and $\boldsymbol{y}_{\text{test}}$ be a test sample and the corresponding class membership vector. Let $\boldsymbol{a}_{\text{test}}$ and $\boldsymbol{b}_{\text{test}}$ be the projections of $\Phi(\boldsymbol{x}_{\text{test}})$ and $\boldsymbol{y}_{\text{test}}$ onto $\boldsymbol{P}_{\Phi(x)}$ and $\boldsymbol{P}_y$, respectively. Then we obtain $\boldsymbol{a}_{\text{test}} = \boldsymbol{P}_{\Phi(x)}^{\mathrm{T}}\ \Phi(\boldsymbol{x}_{\text{test}}) = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}_{\text{test}}$ and $\boldsymbol{b}_{\text{test}} = \boldsymbol{P}_y^{\mathrm{T}}\ \boldsymbol{y}_{\text{test}}$, where $\boldsymbol{K}_{\text{test}} = (\Phi(X))^{\mathrm{T}}\Phi(\boldsymbol{x}_{\text{test}})$ can be computed via the kernel function. According to Eq. (12), we know that $\boldsymbol{y}_{\text{test}}$ can be estimated by

$$\boldsymbol{y}_{\text{test}} = (\boldsymbol{P}_y\boldsymbol{P}_y^{\mathrm{T}})^{-1}\boldsymbol{P}_y\hat{\boldsymbol{P}}\boldsymbol{P}_{\Phi(x)}^{\mathrm{T}}\ \Phi(\boldsymbol{x}_{\text{test}}) = (\boldsymbol{P}_y\boldsymbol{P}_y^{\mathrm{T}})^{-1}\boldsymbol{P}_y\hat{\boldsymbol{P}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}_{\text{test}} \tag{16}$$

where $\hat{\boldsymbol{P}}$ is estimated by the same method as Eq. (10). Moreover, it is notable that the distance metric in Eq. (14) should be replaced by the following formulation:

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(\Phi(\boldsymbol{x}_i) - \Phi(\boldsymbol{x}_j))^{\mathrm{T}}(\Phi(\boldsymbol{x}_i) - \Phi(\boldsymbol{x}_j))} = \sqrt{k(\boldsymbol{x}_i, \boldsymbol{x}_i) - 2k(\boldsymbol{x}_i, \boldsymbol{x}_j) + k(\boldsymbol{x}_j, \boldsymbol{x}_j)} \tag{17}$$

4 Experiments

In this experiment, we use the Japanese Female Facial Expression(JAFFE) database[12-13] and the facial expression image set of Ekman and Friesen[14], respectively, to test the proposed method(fuzzy-KDA/KCCA). The monomial kernel and the Gaussian kernel are respectively used in the fuzzy-KDA/KCCA approach, which are respectively defined as follows:

1) Monomial kernel

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^{\mathrm{T}}\ \boldsymbol{x}_j)^d$$

where d is the degree of the monomial kernel.

2) Gaussian kernel

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{\sigma}\right)$$

where σ is the parameter of Gaussian kernel.

The JAFFE database contains 213 facial images covering all the seven facial expressions(happiness, sadness, surprise, anger, disgust, fear and neutral) posed by 10 Japanese female. Each subject has two to four images for each of the seven expressions. The original images are all sized pixels with a 256-level gray scale. The database of Ekman and Friesen contains 110 images consisting of 6 male and 8 female subjects. Each subject has at most one to two images for each of the seven expressions. The each image consists of pixels with a 256-level gray scale. In the preprocessing stage, we manually locate 34 landmark points from each facial image by referring to the method in Ref. [12]. Fig. 1 illustrates an example of the 34 landmark points.

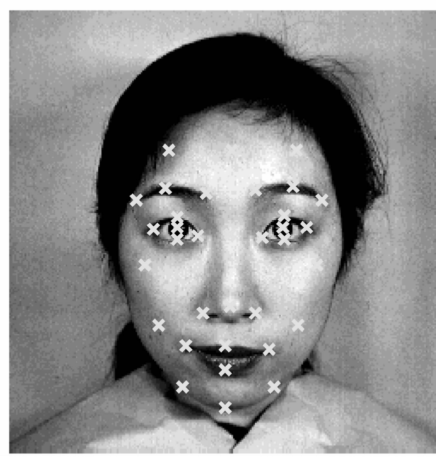

Fig. 1 Sample figure caption

Similar to Ref. [12], after locating the 34 landmark points, we use the Gabor wavelet representation of each facial image at the landmark points to represent the facial features of each image, where all of the wavelet convolution values(magnitudes) at these landmark points are combined into a 1 020 dimensional LG vector. The Gabor kernel is defined as[10-11]

$$\psi_{u,v} = \frac{\|k_{u,v}\|^2}{\sigma^2}\exp\left(-\frac{\|k_{u,v}\|^2\|z\|^2}{2\sigma^2}\right)\cdot \left[\exp(\mathrm{i}k_{u,v}\cdot z) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \tag{18}$$

where u and v represent the orientation and scale of the Gabor kernels, and $k_{u,v}$ is defined as

$$k_{u,v} = k_v\exp(\mathrm{i}\phi_u) \tag{19}$$

where $k_v = \frac{\pi}{2^v}(v \in \{1, 2, \ldots, 5\})$, and $\phi_u = \frac{\pi u}{6}(u \in \{0, 1, 2, \ldots, 5\})$.

Since both databases are relatively small, we only use the "leave-one-class-out" cross validation strategy to conduct this experiment. In the "leave-one-class-out" cross validation strategy, all the images belonging to one subject are used as the testing data and the remaining ones as the training data. The procedure is repeated until all the subjects are used once as the testing data. The recognition rates of the experiment are averaged as the final recognition rate. For comparison purposes, we also conduct the same experiments using the GDA method[9], the LDA method[1], and the KCCA method[11], respectively. Tabs. 1 and 2 show the results of various systems on the JAFFE database and the Ekman and Friesen database, respectively.

Tab. 1 Comparison of average recognition rate on JAFFE facial expression database

Methods	Recognition rate/%
Fuzzy-LDA/CCA	68.31
Fuzzy-KDA/KCCA(Gaussian kernel with $\sigma = 2\times10^6$)	78.69
Fuzzy-KDA/KCCA(Monomial kernel with $d=2$)	71.58
LDA	64.48
GDA(Gaussian kernel with $\sigma = 2\times10^6$)	77.05
GDA(Monomial kernel with $d=2$)	69.95
KCCA[11]	77.08

Tab. 2 Comparison of average recognition rate on Ekman facial expression database

Methods	Recognition rate/%
Fuzzy-LDA/CCA	78.13
Fuzzy-KDA/KCCA(Gaussian kernel with $\sigma = 7\times10^6$)	82.29
Fuzzy-KDA/KCCA(Monomial kernel with $d=2$)	79.17
LDA	76.04
GDA(Gaussian kernel with $\sigma = 7\times10^6$)	78.13
GDA(Monomial kernel with $d=2$)	76.04
KCCA[11]	77.05

From Tabs. 1 and 2, we can see that the proposed method achieves the best performance among the various methods. The reason is that we use the fuzzy membership description rather than the simple binary value membership description for each data point, which is more reasonable for each facial image. Hence, the fuzzy-based facial expression recognition methods can obtain better recognition results. Moreover, from Tabs. 1 and 2 we can also see that the fuzzy-KDA/KCCA method achieves better recognition results than the fuzzy-LDA/CCA. This is because the fuzzy-KDA/KCCA is actually a nonlinear facial feature method, which can extract the nonlinear facial features for the recognition task. In contrast with the fuzzy-KDA/KCCA, the fuzzy-LDA/CCA is only a linear extraction method, which may be less powerful when used for such nonlinear pattern recognition problems as facial expression recognition.

5 Conclusion

This paper proposes a fuzzy linear discriminant analysis method via canonical correlation analysis (fuzzy-LDA/CCA) and applies it to the facial expression recognition task. The major advantage of the fuzzy-LDA/CCA method, compared with the traditional LDA method, is that it provides an effective technique to estimate the degrees of the class membership of an unknown test sample, which is very useful for facial expression analysis since each facial image may contain several emotion categories. We have conducted experiments on two commonly used facial expression databases and showed that the proposed method achieves better performance than the traditional discriminant analysis methods.

References

[1] Duda R O, Hart P E, Stork D G. *Pattern classification* [M]. 2nd ed. New York: Wiley Press, 2000.

[2] Kwak K-C, Pedrycz W. Face recognition using a fuzzy fisherface classifier[J]. *Pattern Recognition*, 2005, **38**(10): 1717 − 1732.

[3] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 711 − 720.

[4] Hardoon D R, Szedmark S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods[J]. *Neural Computation*, 2004, **16**(12): 2639 − 2664.

[5] Barker M, Rayens W. Partial least squares for discrimination [J]. *Journal of Chemometrics*, 2003, **17**(3): 166 − 173.

[6] John S-T, Nello C. *Kernel methods for pattern analysis*[M]. Cambridge: Cambridge University Press, 2004.

[7] Schölkopf B, Smola A J, Müller K-R. Nonlinear component analysis as a kernel eigenvalue problem[J]. *Neural Computation*, 1998, **10**(5): 1299 − 1319.

[8] Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach[J]. *Neural Computation*, 2000, **12**(10): 2385 − 2404.

[9] Zheng Wenming, Zhao Li, Zou Cairong. A modified algorithm for generalized discriminant analysis[J]. *Neural Computation*, 2004, **16**(6): 1283 − 1297.

[10] Wu Qingjiang, Zhou Xiaoyan, Zheng Wenming. Facial expression recognition using fuzzy kernel discriminant analysis [C]//*Lecture Notes in Computer Science*: *Fuzzy System and Knowledge Discovery*. Berlin-Heidelberg: Springer-Verlag, 2006: 780 − 783.

[11] Zheng Wenming, Zhou Xiaoyan, Zou Cairong, et al. Facial expression recognition using kernel canonical correlation analysis (KCCA) [J]. *IEEE Transactions on Neural Networks*, 2006, **17**(1): 233 − 238.

[12] Lyons M, Budynek J, Akamatsu S. Automatic classification of single facial images[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1999, **21**(12): 1357 − 1362.

[13] Zhang Z, Lyons M, Schuster M, et al. Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perception[C]//*Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*. Nara, Japan, 1998: 454 − 459.

[14] Ekman P, Friesen W V. Pictures of facial affect [R]. San Francisco: Human Interaction Laboratory of University of California Medical Center, 1976.

基于模糊 LDA/CCA 的面部表情识别

周晓彦[1,2]　郑文明[3]　邹采荣[1,4]　赵　力[1]

([1] 东南大学信息科学与工程学院,南京 210096)
([2] 南京信息工程大学电子与信息工程学院,南京 210044)
([3] 东南大学学习科学中心,南京 210096)
([4] 佛山科技学院,佛山 528000)

摘要:提出了一种新颖的基于典型相关分析(CCA)的模糊判别分析方法(fuzzy-LDA/CCA),并应用于面部表情识别问题. 首先为每幅表情图像建立一个相关联的类模糊隶属度矢量,用于表示表情图像与基本表情类别的隶属关系,在此基础上应用 CCA 方法建立表情图像同表情类别的关系表达式,最后通过对表情图像的类隶属度矢量的估计来实现表情的分类. 此外,还将 fuzzy-LDA/CCA 方法在核空间中进行了非线性推广,从而来解决非线性判别分析的问题. 实验证明提出的方法获得了更好的识别效果.

关键词:模糊判别分析;典型相关分析;面部表情识别

中图分类号:TP391.41