Neural Comput & Applic
DOI 10.1007/s00521-011-0528-3

Weighted linear embedding: utilizing local and nonlocal information sufficiently

Jun Yin · Jingbo Zhou · Zhong Jin · Jian Yang

Received: 27 October 2010 / Accepted: 17 January 2011

Abstract Local discriminant embedding (LDE) only utilizes the local information and ignores the nonlocal information. Although linear discriminant analysis (LDA) utilizes the local information and the nonlocal information simultaneously, it treats these two kinds of information equally. As we know, the local information and the nonlocal information are both effective for feature extraction, but they have different roles in feature extraction. To utilize the local information and the nonlocal information simultaneously and utilize them distinctively, a new feature extraction approach called weighted linear embedding (WLE) is proposed by using the Gaussian weighting. Further, a method to set the optimal parameter of the Gaussian weighting is put forward. WLE is evaluated on YALE, FERET face databases, the PolyU palmprint database, and the PolyU finger-knuckle-print database. The experimental results demonstrate the effectiveness of WLE.

Keywords Feature extraction · Principal components analysis · Linear discriminant analysis · Local discriminant embedding

1 Introduction

Since there are large volumes of high-dimensional data in numerous real-world applications, dimensionality reduction is a fundamental problem in many scientific fields such as visualization, computer vision, and pattern recognition. Feature extraction is an essential technique for dimensionality reduction. With respect to pattern recognition, feature extraction is an effective approach to reveal the distinctive features from the original data [1].

Principal components analysis (PCA) [2] and linear discriminant analysis (LDA) [3] are two of the most popular feature extraction methods. They project the data into a low-dimensional subspace via a projection matrix. These two methods are based upon the assumption that the data are Gaussian distributed and the testing data are drawn from the same distribution as the training data. However, this assumption cannot always be satisfied, because of the limitation of data collection and the variations of input.

Recently, manifold learning methods such as Isomap [4], locally linear embedding (LLE) [5], and Laplacian eigenmap (LE) [6] are proposed for feature extraction. They seek to find the intrinsic low-dimensional nonlinear data structure hidden in observation space. However, because it is difficult to map the new testing data to the low-dimensional space, these manifold learning algorithms cannot be easily extended for classification. Locality preserving projections (LPP) [7], locally embedded analysis (LEA) [8], and unsupervised discriminant projection (UDP) [9] address this difficulty by finding an explicit mapping for all the data. They are linear manifold learning methods that consider the linear manifold structure of the data. Nevertheless, these algorithms only focus on preserving the localities and similarities of data, but ignore the class information.

To further improve the discriminative power of manifold learning methods, some new discriminant analysis methods such as marginal Fisher analysis (MFA) [10], local discriminant embedding (LDE) [11], conformal embedding analysis (CEA) [12], locality sensitive discriminant analysis (LSDA) [13], neighborhood discriminant projection (NDP) [14], Laplacian bidirectional maximum margin criterion (LBMMC) [26], and multi-manifold discriminant analysis

J. Yin (✉) · J. Zhou · Z. Jin · J. Yang
School of Computer Science and Technology,
Nanjing University of Science and Technology,
210094 Nanjing, China
e-mail: yinjun8429@163.com

Published online: 02 February 2011

(MMDA) [28] combine Fisher criterion [3] with manifold learning criterion. They consider the local information and the class information simultaneously, but neglect the nonlocal information between samples which are far from each other. However, the nonlocal information may be useful for classification [9].

In this paper, we propose a new feature extraction approach called weighted linear embedding (WLE), in which the local information, the nonlocal information, and the class information are all used sufficiently. For WLE, we make use of the local information and the nonlocal information distinctively according to their different effects. We highlight some advantages of WLE algorithm as follows:

1. WLE is a supervised learning method, which leads to its better classification ability than unsupervised learning method like PCA and LPP.
2. By manifold learning theory, we know that the local information is more important than the nonlocal information in discovering the underlying manifold structure of the dataset [5, 6]. WLE treats the local information and the nonlocal information differently, so it has better performance than LDA that treats these two kinds of information equally.
3. WLE considers the nonlocal information and the local information simultaneously. So it has better discriminative power than the methods which only use the local information like LDE.

In the past few decades, biometrics techniques such as face recognition [2, 3, 15, 27], palmprint recognition [16–18], and finger-knuckle-print (FKP) recognition [19–21] have become the hot topics of pattern recognition and computer vision. In this paper, WLE is used for extracting the feature of face, palmprint, and finger-knuckle-print, and the extracted feature is used for recognition. We do recognition experiments on YALE, FERET face databases, the PolyU palmprint database, and the PolyU finger-knuckle-print database. The experimental results demonstrate the superior characteristics of WLE.

The rest of the paper is organized as follows: Sect. 2 outlines LDA and LDE. Section 3 describes the idea and the formulation of WLE in detail. Section 4 presents the experiments. Section 5 gives our conclusions.

For convenience, the important notations used throughout the rest of the paper are listed in Table 1.

2 LDA and LDE

2.1 Linear discriminant analysis (LDA)

LDA aims to find a projection vector such that the ratio of between-class scatter to the within-class scatter is maximized after the projection. The between-class scatter S_b and the within-class scatter S_w are denoted as

$$S_b = \frac{1}{M}\sum_{i=1}^{c} l_i (m_i - m)(m_i - m)^T \tag{1}$$

and

$$S_w = \frac{1}{M}\sum_{i=1}^{c}\sum_{j=1}^{l_i} (x_{ij} - m_i)(x_{ij} - m_i)^T, \tag{2}$$

where x_{ij} is the jth training sample in class i. The criterion of LDA is maximizing

$$J_F(w) = \frac{J_b(w)}{J_w(w)} = \frac{w^T S_b w}{w^T S_w w}. \tag{3}$$

This can be transformed to solve the generalized eigenvector corresponding to the largest eigenvalue in $S_b w = \lambda S_w w$. Then, we could obtain the projection matrix of LDA by selecting d eigenvectors of $S_w^{-1} S_b$ corresponding to the d largest nonzero eigenvalues.

There is another way to describe LDA. First, we define an adjacency matrix H by

$$H(i,j) = \begin{cases} 1, & \text{if } x_i \text{ and } x_j \text{ belong to the same class} \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

Then, S_b and S_w are defined as

$$S_b = \frac{1}{2}\frac{1}{MM}\sum_{i=1}^{M}\sum_{j=1}^{M} (1 - H(i,j))(x_i - x_j)(x_i - x_j)^T \tag{5}$$

and

$$S_w = \frac{1}{2}\frac{1}{MM}\sum_{i=1}^{M}\sum_{j=1}^{M} H(i,j)(x_i - x_j)(x_i - x_j)^T. \tag{6}$$

Yang et al. [9] have proved that using S_b and S_w defined by (5) and (6) to perform LDA is equivalent to the traditional LDA when each class has the same number of training samples.

Table 1 Summary of the notations used

Notations	Description
n	Sample dimension
c	Number of classes
M	Number of the training samples
l_i	Number of training samples in class i
x_i	The ith training sample
m_i	Mean of the training samples in class i
m	Mean of all training samples

2.2 Local discriminant embedding (LDE)

LDE combines the local information and the class label information of samples. It seeks to keep neighboring samples close if they are from the same class and prevent the samples of other classes from entering the neighborhood.

LDE constructs neighbor graphs G_w for the samples from the same class and G_b for the samples from different classes. For G_w, an edge is added between x_i and x_j, if x_i is one of x_j's k_w-nearest neighbors, and they are from the same class. For G_b, an edge is added between x_i and x_j, if x_i is one of x_j's k_b-nearest neighbors, and they are from different classes.

Then, the adjacency matrices H_w for G_w and H_b for G_b are defined by

$$H_w(i,j) = \begin{cases} 1, & \text{if } x_i \text{ and } x_j \text{ are connected in } G_w \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

and

$$H_b(i,j) = \begin{cases} 1, & \text{if } x_i \text{ and } x_j \text{ are connected in } G_b \\ 0, & \text{otherwise} \end{cases}. \tag{8}$$

The Laplacian matrix pair [6] is computed like LPP [7] by

$$L_w = D_w - H_w \tag{9}$$

and

$$L_b = D_b - H_b, \tag{10}$$

where D_w and D_b are diagonal matrices with diagonal elements $D_w[i,i] = \sum_{j=1}^{M} H_w(i,j) \quad (i = 1,2,\ldots,M)$ and $D_b[i,i] = \sum_{j=1}^{M} H_b(i,j) \quad (i = 1,2,\ldots,M)$. The criterion of LDE is maximizing

$$J_{LDE}(w) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M} H_b(i,j)(w^T x_i - w^T x_j)^2}{\sum_{i=1}^{M}\sum_{j=1}^{M} H_w(i,j)(w^T x_i - w^T x_j)^2}. \tag{11}$$

This can be transformed to

$$J_{LDE}(w) = \frac{w^T X L_b X^T w}{w^T X L_w X^T w}, \tag{12}$$

where $X = (x_1, x_2, \ldots, x_M)$ is the data matrix. So the projection matrix of LDE could be constructed by selecting d eigenvectors of $XL_bX^Tw = \lambda XL_wX^Tw$ corresponding to the d largest nonzero eigenvalues.

3 Weighted linear embedding

From (5) and (6), we can see that the local neighbor information and the nonlocal information of each sample are treated equally in LDA when constructing S_b and S_w. However, according to the manifold learning theory, we know that the local information may be more important for discovering the underlying manifold structure of the dataset [5, 6]. So the local information and the nonlocal information should be treated differently. As discussed in Sect. 2.2, only the local information of each sample is considered, while the nonlocal information is neglected in LDE. Nevertheless, the nonlocal information may be also useful for feature extraction and classification [9]. Hence, both the local information and the nonlocal information should be taken into account.

3.1 Idea

To consider the local information and the nonlocal information together and treat them differently, we make use of the Gaussian weighting. The Gaussian weighting function is defined by

$$\exp\left(-\frac{\|x_i - x_j\|^2}{t}\right). \tag{13}$$

The purpose of the weighting is to indicate the importance of the relation between x_i and x_j. Gaussian function is a strictly monotone decreasing function with respect to the distance between x_i and x_j. Smaller is the distance, more important is the relation. So this exactly conforms to the fact that the local information is more important than the nonlocal information.

Here, two adjacency matrices $\tilde{H}_b$ and $\tilde{H}_w$ using the Gaussian weighting are defined by

$$\tilde{H}_b(i,j) = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{t_b}\right), & \text{if } x_i \text{ and } x_j \text{ belong to different classes} \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

and

$$\tilde{H}_w(i,j) = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{t_w}\right), & \text{if } x_i \text{ and } x_j \text{ belong to the same class} \\ 0, & \text{otherwise} \end{cases}. \tag{15}$$

WLE aims to find a projection vector such that the ratio of the weighted between-class scatter to the weighted within-class scatter is maximized after the projection. The weighted between-class scatter and the weighted within-class scatter after the projection w are defined by

$$\tilde{J}_b(w) = \sum_{i=1}^{M}\sum_{j=1}^{M} \tilde{H}_b(i,j)(w^T x_i - w^T x_j)^2 \tag{16}$$

and

$$\tilde{J}_w(w) = \sum_{i=1}^{M}\sum_{j=1}^{M}\tilde{H}_w(i,j)(w^Tx_i - w^Tx_j)^2. \quad (17)$$

This can be transformed to

$$\tilde{J}_b(w) = w^TX(\tilde{D}_b - \tilde{H}_b)X^Tw \quad (18)$$

and

$$\tilde{J}_w(w) = w^TX(\tilde{D}_w - \tilde{H}_w)X^Tw, \quad (19)$$

where $\tilde{D}_w$ and $\tilde{D}_b$ are diagonal matrices with diagonal elements $\tilde{D}_w[i,i] = \sum_{j=1}^{M}\tilde{H}_w(i,j) \quad (i = 1,2,\ldots,M)$ and $\tilde{D}_b[i,i] = \sum_{j=1}^{M}\tilde{H}_b(i,j) \quad (i = 1,2,\ldots,M)$.Here, we also define Laplacian matrix pair by

$$\tilde{L}_w = \tilde{D}_w - \tilde{H}_w \quad (20)$$

and

$$\tilde{L}_b = \tilde{D}_b - \tilde{H}_b. \quad (21)$$

Now, the criterion of WLE is maximizing

$$J_{WLE}(w) = \frac{\tilde{J}_b(w)}{\tilde{J}_w(w)} = \frac{w^TX\tilde{L}_bX^Tw}{w^TX\tilde{L}_wX^Tw}. \quad (22)$$

We could get the projection matrix of WLE by selecting d eigenvectors of $X\tilde{L}_bX^Tw = \lambda X\tilde{L}_wX^Tw$ corresponding to the d largest nonzero eigenvalues.

3.2 The method to set the parameter

Parameter selection occupies a significant position in the algorithm. For the proposed WLE algorithm, the parameters t_b and t_w appear in weighting calculation. The weighting that could reflect the difference between the local information and the nonlocal information should be selected. In addition, to make the nonlocal information effective, the weighting of the nonlocal sample pair cannot be too small compared with that of the local sample pair. Based on these two points, we give an approach to set t_b and t_w.

Now, we discuss how to select t_w. Compute all the distances between every training sample pair from the same class. Assume that d_L and d_N are any two distances, $d_N > d_L$. Here, d_N can be seen as a distance of the nonlocal sample pair, and d_L can be seen as a distance of the local sample pair. The Gaussian weighting constructed by d_L and d_N is

$$g_L(t_w) = \exp\left(-\frac{d_L^2}{t_w}\right) \quad (23)$$

and

$$g_N(t_w) = \exp\left(-\frac{d_N^2}{t_w}\right). \quad (24)$$

The difference of $g_L(t_w)$ and $g_N(t_w)$ could be represented by the ratio of $g_L(t_w)$ to $g_N(t_w)$:

$$R = \frac{g_L(t_w)}{g_N(t_w)} = \exp\left(\frac{d_N^2 - d_L^2}{t_w}\right) = \exp\left(\frac{(d_N + d_L)(d_N - d_L)}{t_w}\right). \quad (25)$$

The expectation of $d_N + d_L$ can be estimated by $2\bar{d}_w$ and $d_N - d_L$ by $2\Delta_w$ approximately, where $\bar{d}_w$ is the mean of all the distances between every training sample pair from the same class, and Δ_w is the standard deviation. Now, R can be estimated by $\exp\left(\frac{4\bar{d}_w\Delta_w}{t_w}\right)$, and it is only decided by $p_w = \frac{4\bar{d}_w\Delta_w}{t_w}$. We should select an appropriate p_w. If p_w is given, t_w can be computed by

$$t_w = \frac{4\bar{d}_w\Delta_w}{p_w}. \quad (26)$$

In the same way, if the appropriate parameter p_b is given, t_b can be computed by

$$t_b = \frac{4\bar{d}_b\Delta_b}{p_b}, \quad (27)$$

where $\bar{d}_b$ is the mean of all the distances between every training sample pair from different classes, and Δ_b is the standard deviation.

Based on the above analysis, now WLE only needs to set the parameters p_w and p_b. p_w and p_b are much easier to set than t_w and t_b. The optimal range of them is not affected by the database very much. Through experiments, it is found that WLE has the best performance when p_w is within the interval (0, 2) and p_b is (0, 8).

3.3 Links to LDA and LDE

Here, we discuss the connection of WLE and LDA and the connection of WLE and LDE.

We know any positive value can be selected for p_w and p_b. On one hand, when we set $p_w = 0$ and $p_b = 0$, according to (26) and (27), we will have $t_w = +\infty$ and $t_b = +\infty$. In this case, all the weighting $\exp\left(-\frac{\|x_i - x_j\|^2}{t_w}\right)$ and $\exp\left(-\frac{\|x_i - x_j\|^2}{t_b}\right)$ will be equal to one. Here, WLE is equal to LDA in fact.

On the other hand, when we make p_w and p_b adequately big, t_w and t_b will become very small. Let us see training sample pair from the same class. When t_w is very small, according to (25), the ratio of the weighting of the local sample pair $g_L(t_w)$ to the weighting of the nonlocal sample pair $g_N(t_w)$ will be greatly big. If this happens, $g_N(t_w)$ can be seen as zero compared with $g_L(t_w)$. Now, for training sample pair from the same class, only the local information is considered actually. Similar to training sample pair from the same class, when t_b is very small, for training sample pair from different classes, the local information is only considered too. So WLE is similar to LDE when p_w and p_b are adequately big.

Therefore, LDA and LDE can be seen as two special cases of WLE.

3.4 The algorithm of WLE

As mentioned in Sect. 3.1, WLE can be transformed to generalized eigenvalue problem: $X\tilde{L}_bX^Tw = \lambda X\tilde{L}_wX^Tw$. However, $X\tilde{L}_wX^T$ is often singular, because the training sample size is smaller than the dimension of the data space. This is so-called small sample size (SSS) problem [22]. To address this issue, we first use PCA to reduce the dimension of the data so that $X\tilde{L}_wX^T$ is nonsingular in the PCA subspace.

The algorithm of WLE can be summarized as follows:

Step 1 Perform PCA on the original data. Denote the transformation matrix of PCA by W_{PCA}. Then, the new PCA feature $x_i^{PCA} = P^Tx_i \ (i = 1, 2, \ldots, M)$

Step 2: Use the new PCA feature $x_i^{PCA} \ (i = 1, 2, \ldots, M)$ to construct the adjacency matrices $\tilde{H}_b$ and $\tilde{H}_w$ by (14) and (15). Then, compute $\tilde{J}_b(w)$ and $\tilde{J}_w(w)$ using the adjacency matrices by (18) and (19)

Step 3: Solve the criterion of WLE (22) by calculating the eigenvectors $q_1, q_2, \ldots, q_d$ of $(X\tilde{L}_wX^T)^{-1}X\tilde{L}_bX^T$ corresponding to the d largest nonzero eigenvalues

Step 4: Let $Q = (q_1, q_2, \ldots, q_d)$. The projection matrix of WLE is $W_{WLE} = PQ$. The new WLE feature of a sample x could be formed by

$$y = W_{WLE}^T x \tag{28}$$

4 Experiments

In this section, the recognition experiments on YALE and FERET face databases, the PolyU palmprint database, and the PolyU finger-knuckle-print database are designed to evaluate the performance of WLE. WLE algorithm is compared with PCA, LPP, LDA, and LDE. The optimal parameters for LPP, LDE, and WLE are set. For classification, the nearest-neighbor classifier is employed.

4.1 Experiment on YALE database

The YALE face database was constructed at the YALE Center for Computational Vision and Control. It contains 165 gray-scale images of 15 individuals. The images demonstrate variations in lighting, facial expression, and with/without glasses. In our experiment, every image is manually cropped and resized to 100 × 80 pixels. Figure 1 shows eleven images of one person.

In the first experiment, WLE is performed with different p_w and p_b. We use the first four, five, and six images per person for training and the rest images for testing. Firstly, we fix p_b at 2 and vary p_w from 0 to 8 with a step 0.4. The recognition rates of WLE versus the variation of p_w are shown in Fig. 2. From Fig. 2, we can see that the maximal recognition rates of WLE are mainly received when p_w is within the interval (0, 2), no matter how many samples are used for training. Secondly, the parameter p_w is fixed at 2, and the parameter p_b is varied from 0 to 20 with a step 1. The recognition rates of WLE versus the variation of p_b are

Fig. 1 Sample images of one person on YALE database

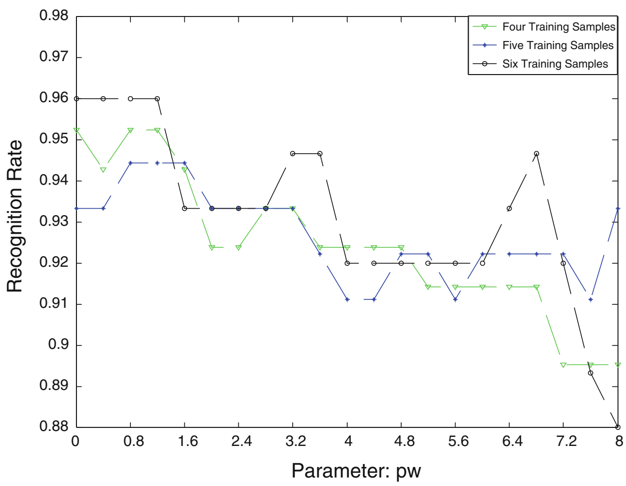

Fig. 2 The recognition rates of WLE versus the variation of p_w on YALE database

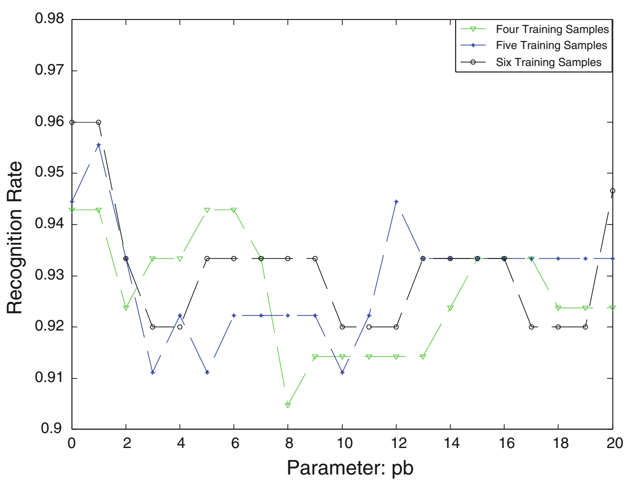

Fig. 3 The recognition rates of WLE versus the variation of p_b on YALE database

shown in Fig. 3. From Fig. 3, we find that WLE has a good performance when p_b is within the interval (0, 8), irrespective of training samples. These experimental results demonstrate the conclusion of Sect. 3.3. From Figs. 2 and 3, we can also see that the recognition rates of WLE versus the variation of p_w or p_b fluctuate. This phenomenon can be interpreted as follows: With the increase of p_w or p_b, the effect of the local information is enhanced. However, meanwhile, the effect of the nonlocal information is reduced. These two aspects cause the fluctuation of the recognition rates together.

In the second experiment, we compare the performance of PCA, LPP, LDA, LDE, and WLE. Four images of each subject are randomly selected for training, and the remaining images form the testing sample set. The experiment is repeated for 20 times. The maximal average recognition rates and standard deviations of each method and the corresponding dimensions and optimal parameters are shown in Table 2, where k represents k-nearest-neighbor parameter of LPP. From Table 2, it can be seen that supervised algorithms LDA, LDE, and WLE outperform unsupervised algorithms PCA and LPP. Besides, LDE performs a little better than LDA. We can also see that WLE performs better than LDA and LDE, and it reaches the maximal recognition rate when p_w is equal to 1 and p_b is 0.4.

4.2 Experiment on FERET database

The FERET face database was sponsored by the US Department of Defense through the DARPA Program [23, 24]. It has become a standard database for testing and evaluating face recognition algorithms. We do experiment on a subset of the FERET database. The subset is composed of 1,400 images of 200 individuals, each individual has seven images. It involves variations in face expression, pose, and illumination. In the experiment, the facial portion of the original image was cropped based on the location of eyes and mouth. Then, we resized the cropped images to 80 × 80 pixels and preprocess them by histogram equalization. Seven sample images of one individual are show in Fig. 4.

The first three images per person are chosen for training, and the remaining four images for testing. PCA, LPP, LDA, LDE, and WLE are applied for feature extraction. Table 3 shows the best recognition results and the corresponding dimensions and optimal parameters. Fig. 5 shows the recognition rates of each method versus the variation of dimensions. From Table 3 and Fig. 5, three main points can be seen. First, LDA, LDE, and WLE outperform PCA

Table 2 The maximal average recognition rates (percent) and standard deviation (percent) of PCA, LPP, LDA, LDE, and WLE on YALE database across 20 runs and the corresponding dimensions and parameters

Method	PCA	LPP	LDA	LDE	WLE
Dimension	19	18	13	14	16
Average recognition rate	90.4	91.9	92.8	93.1	94.5
Standard deviation	2.4	3.0	2.9	3.8	3.0
Parameter	–	$k = 2$	–	$k_w = 2, k_b = 4$	$p_w = 1, p_b = 0.4$

Fig. 4 Sample images of one person on FERET database

Table 3 The maximal recognition rates (percent) of PCA, LPP, LDA, LDE, and WLE on a subset of FERET database and the corresponding dimensions and parameters

Method	PCA	LPP	LDA	LDE	WLE
Dimension	151	150	51	73	65
Recognition rate	54.1	45.1	54.9	56.9	59.1
Parameter	–	$k = 2$	–	$k_w = 1$, $k_b = 2$	$p_w = 1$, $p_b = 7$

Table 4 The maximal recognition rates (percent) of PCA, LPP, LDA, LDE, and WLE on the PolyU FKP database and the corresponding dimensions and parameters

Method	PCA	LPP	LDA	LDE	WLE
Dimension	95	71	60	53	48
Recognition rate	59.0	58.5	64.5	64.0	66.5
Parameter	–	$k = 3$	–	$k_w = 3$, $k_b = 9$	$k_w = 1.4$, $k_b = 0$

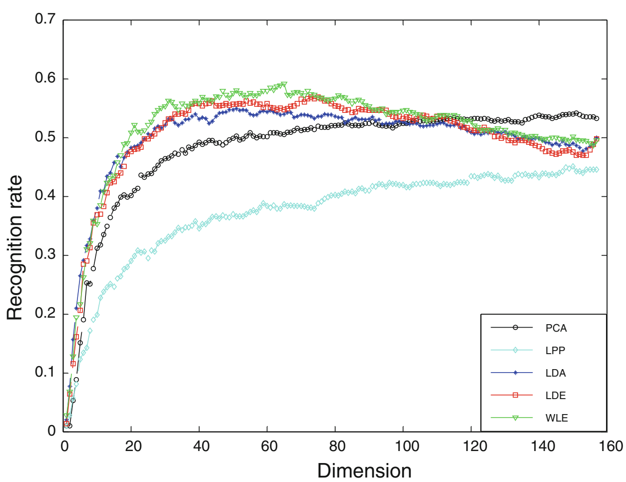

Fig. 5 The recognition rates of PCA, LPP, LDA, LDE, and WLE versus the dimensions on a subset of FERET database

and LPP. Second, LDE performs better than LDA. Third, WLE performs best, and the optimal parameters for WLE are $p_w = 1$ and $p_b = 7$.

4.3 Experiment on the PolyU FKP database

FKP images on the PolyU FKP database were collected from 165 volunteers. These images are collected in two separate sessions. In each session, the subject was asked to provide six images for each of the left index finger, the left middle finger, the right index finger, and the right middle finger. The images were processed by ROI extraction algorithm described in [20]. In the experiment, we select 1,200 FKP images of the right index finger of 100 subjects. These selected images were resized to 55×110 pixels and preprocessed by histogram equalization. Figure 6 shows twelve sample images of one right index finger.

The first six FKP images collected in the first session are used for training and the rest six collected in the second session for testing. PCA, LPP, LDA, LDE, and WLE are performed for feature extraction separately. Table 4 lists the maximal recognition rates of five methods and the corresponding dimensions and optimal parameters. It can be seen from Table 4 that LDA outperforms LDE in this experiment. This result is inconsistent with the results on YALE and FERET face databases. However, we can also see a consistent result that WLE has the best performance among five methods. It has been analyzed that LDA and LDE could be seen as two special cases of WLE. Therefore, if the optimal parameters are selected, WLE can always outperform LDA and LDE, no matter which one of them performs better.

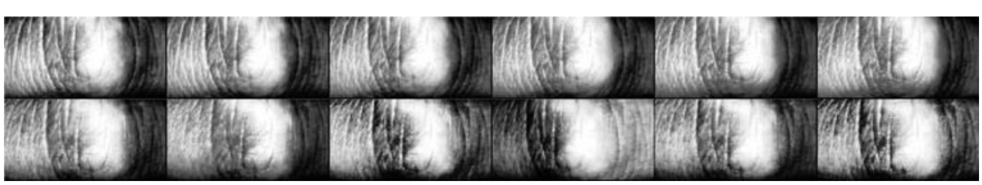

Fig. 6 Sample images of one right index finger

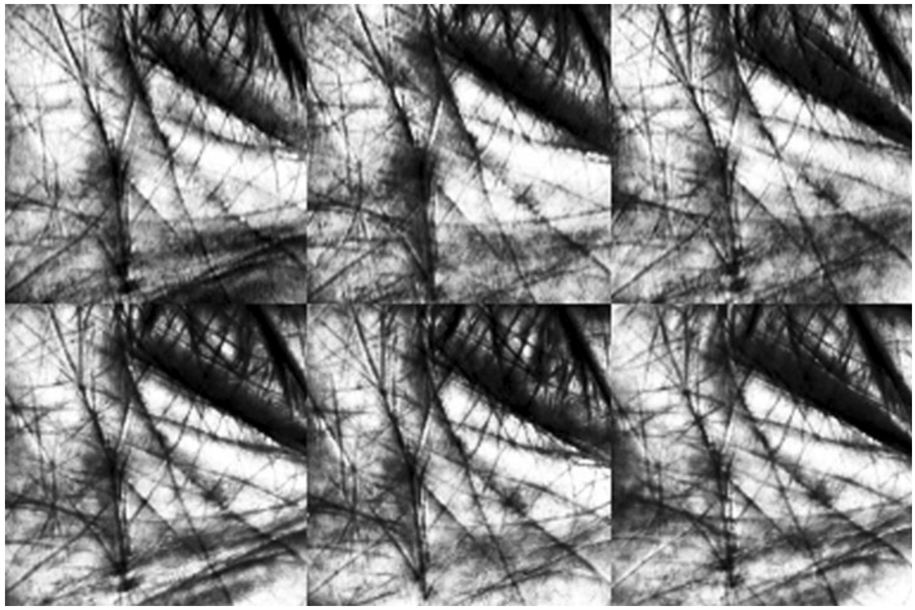

Fig. 7 Sample images of one palm

4.4 Experiment on the PolyU palmprint database

We do experiment on a subset of the the PolyU palmprint database. The subset contains 600 images of 100 different palms with six samples for each palm. 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 central part of each original image was automatically cropped using the algorithm mentioned in [25]. The cropped images were resized to 128 × 128 pixels and preprocessed by histogram equalization. Figure 7 shows six sample images of one palm.

The first three palmprint images captured in the first session are chosen for training and the remaining three captured in the second session for testing. PCA, LPP, LDA, LDE, and WLE are used for feature extraction. Table 5 lists the best recognition results of five algorithms and the corresponding dimensions and optimal parameters. From Table 5, it can be seen that LPP has a good performance in this database, and it outperforms PCA, LDA, and LDE. However, WLE still has the best performance. The maximal recognition rate of WLE reaches 93.3% when p_w is equal to 0.4 and p_b is 1.

Table 5 The maximal recognition rates (percent) of PCA, LDA, LDE, and WLE on the PolyU palmprint database and the corresponding dimensions and parameters

Method	PCA	LPP	LDA	LDE	WLE
Dimension	105	97	94	93	95
Recognition rate	88.0	92.0	91.7	88.3	93.3
Parameter	–	$k = 2$	–	$k_w = 1$, $k_b = 17$	$p_w = 0.4$, $p_b = 1$

5 Conclusions and future work

In this paper, we developed a new algorithm called weighted linear embedding (WLE) for feature extraction. Unlike LDA and LDE, WLE considers the local information, the nonlocal information, and the class information together, and the local information and the nonlocal information are treated differently through using the weighting. A method to set the parameters of WLE is put forward, and the optimal range of the parameters is given. The experimental results on YALE and FERET face databases, the PolyU palmprint database, and the PolyU finger-knuckle-print database indicate the effectiveness of WLE.

The Gaussian function is selected as the weighting in this paper. If there are more effective weighting that can represent the different importance of the local information and the nonlocal information more exactly, this weighting may improve the recognition result further. We will investigate this problem in the future.

Acknowledgments This project is supported by the National Science Foundation of China under Grants No. 60632050, No. 60873151, No. 60973098, No. 60705006.

References

1. Duda RO, Hart PE, Stork DG (2001) Pattern classification. Wiley-Interscience, New York

2. Turk M, Pentland AP (1991) Face recognition using eigenfaces. In: Proceedings of IEEE conference on computer vision and pattern recognition, Maui, pp 586–591
3. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720
4. Tenenbaum JB, DeSilva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323
5. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326
6. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 15(6):1373–1396
7. He XF, Yan SC, Hu YX, Niyogi P, Zhang HJ (2005) Face recognition using Laplacianfaces. IEEE Trans Pattern Anal Mach Intell 27(3):328–340
8. Fu Y, Huang TS (2005) Locally linear embedded eigenspace analysis Univ. Illinois Urbana-Champaign, Urbana, IL, Tech. Rep. IFP-TR, ECE
9. Yang J, Zhang D, Yang JY, Niu B (2007) Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics. IEEE Trans Pattern Anal Mach Intell 29(4):650–664
10. Yan SC, Xu D, Zhang B, Zhang HJ, Yang Q, Lin S (2007) Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell 29(1):40–51
11. Chen HT, Chang HW, Liu TL (2005) Local discriminant embedding and its variants. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 846–853
12. Fu Y, Liu M, Huang TS (2007) Conformal embedding analysis with local graph modeling on the unit hypersphere. In: Proceedings of IEEE conference on computer vision and pattern recognition, Minneapolis, MN, pp 1–6
13. Cai D, He X, Zhou K, Han J, Bao H (2007) Locality sensitive discriminant analysis. In: Proceedings of international joint conference on artificial intelligence, Hyderabad, India, pp 708–713
14. You Q, Zheng N, Du S, Wu Y (2006) Neighborhood discriminant projection for face recognition. In: Proceedings of IEEE international conference on pattern recognition, Hong Kong, pp 532–535
15. Zhang L, Yang M, Feng ZZ, Zhang D (2010) On the dimensionality reduction for sparse representation based face recognition. In: Proceedings of IEEE international conference on pattern recognition, Istanbul, Turkey, pp 1237–1240
16. Zhang L, Guo ZH, Wang Z, Zhang D (2007) Palmprint verification using complex wavelet transform. In: Proceedings of IEEE international conference on image processing, San Antonio, Texas, USA, pp 417–420
17. Zhang L, Zhang D (2004) Characterization of palmprints by wavelet signatures via directional context modeling. IEEE Trans System, Man Cybern B 34(3):1335–1347
18. Zhang D, Lu G, Li W, Zhang L, Luo N (2009) Palmprint recognition using 3-D information. IEEE Trans System Man Cybern C 39(5):505–519
19. Zhang L, Zhang L, Zhang D (2009) Finger-knuckle-print: a new biometric identifier. In: Proceedings of IEEE international conference on image processing, Cairo, pp 1981–1984
20. Zhang L, Zhang L, Zhang D, Zhu HL (2010) On-line finger-knuckle-print verification for personal authentication. Pattern Recogn 43(7):2560–2571
21. Zhang L, Zhang L, Zhang D, Zhu HL (2010) Ensemble of local and global information for finger-knuckle-print recognition. Pattern Recogn (in press)
22. Fukunaga K (1990) Introduction to statistical pattern recognition. Academic Press, New York
23. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The feret evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 22(10):1090–1104
24. Phillips PJ, Wechsler H, Huang J, Rauss P (1998) The feret database and evaluation procedure for face recognition algorithms. Image Vis Comput 16(5):295–306
25. Zhang D (2004) Palmprint authentication, Kluwer Academic
26. Yang WK, Wang JG, Ren MW, Yang JY, Zhang L, Liu GH (2009) Feature extraction based on Laplacian bidirectional maximum margin criterion. Pattern Recogn 42(11):2327–2334
27. Yang M, Zhang L (2010) Gabor feature based sparse representation for face recognition with Gabor Occlusion Dictionary. In: Proceedings of European conference on computer vision, Crete, Greece
28. Yang WK, Sun CY, Zhang L (2010) Face recognition using a multi-manifold discriminant analysis method. In: Proceedings of IEEE international conference on pattern recognition, Istanbul, Turkey, pp 527–530