Optical Engineering

SPIEDigitalLibrary.org/oe

Dimensionality reduction via locally reconstructive patch alignment

Yi Chen
Jun Yin
Jie Zhu
Zhong Jin

Dimensionality reduction via locally reconstructive patch alignment

Yi Chen
Nanjing University of Science and Technology
School of Computer Science and Technology
Nanjing 210094, China
E-mail: cystory@qq.com

Jun Yin
Shanghai Maritime University
College of Information Engineering
Shanghai 201306, China

Jie Zhu
Nanjing Xiaozhuang University
School of Mathematics & Information Technology
Nanjing 211171, China

Zhong Jin
Nanjing University of Science and Technology
School of Computer Science and Technology
Nanjing 210094, China

Abstract. Based on the local patch concept, we proposed locally reconstructive patch alignment (LRPA) for dimensionality reduction. For each patch, LRPA aims to find the low-dimensional subspace in which the reconstruction error of the within-class nearest neighbors is minimized and the reconstruction error of the between-class nearest neighbors is maximized. LRPA preserves the local structure hidden in the high-dimensional space. More importantly, LRPA has natural connections with linear regression classification (LRC). While LRC uses reconstruction errors as the classification rule, a sample can be classified correctly when the within-class reconstruction error is minimal. The goal of LRPA makes it cooperate well with LRC. The experimental results on the extended Yale B (YALE-B), AR, PolyU finger knuckle print, and the palm print databases demonstrate LRPA plus LRC is an effective and robust pattern-recognition system. [DOI: 10.1117/1.OE.51.7.077208]

Subject terms: dimensionality reduction; patch alignment; discriminative locality alignment; face recognition; finger knuckle-print recognition.

Paper 120093 received Jan. 19, 2012; revised manuscript received Jun. 7, 2012; accepted for publication Jun. 27, 2012; published online Aug. 3, 2012.

1 Introduction

Nowadays, dimensionality reduction (DR) techniques have drawn considerable attention in image processing and pattern recognition fields. DR is to find a meaningful low-dimensional representation of high dimensional data. With respect to pattern recognition, DR is an effective way to overcome the "curse of dimensionality."[1] And, more importantly, it reveals the distinctive features from the original data for pattern matching.

Among the existing DR techniques, principle component analysis (PCA)[2] and linear discriminant analysis (LDA)[3] are two well-known tools in the appearance-based approaches. While PCA achieves simply object reconstruction, LDA optimizes the low-dimensional representation of the objects with focus on the most discriminant feature extraction. It is generally believed that, when it comes to solving problems of pattern classification, LDA-based algorithms outperform PCA based ones.[3,4] Unfortunately, it has been pointed out that drawbacks still exist in LDA. For example, LDA suffers from the small sample size problem when the within-class scatter matrix is singular. Moreover, LDA can only extract at most $c-1$ features (c is the number of total classes), which is suboptimal for many applications. In the past decades, numerous LDA variants[5–10] have been developed to solve the limitations of LDA.

Despite of the drawbacks mentioned above, LDA and its extensions fail to discover the nonlinear structure hidden in the high-dimensional non-Gaussian distributed data, since they are based on global Euclidean structure. Many nonlinear techniques[11–18] have been proposed to discover the nonlinear structure of the manifold. Among them, locality preserving projections[14] and neighborhood preserving embedding (NPE)[13] are two widely used tools for DR. They aim to preserving the local information in the low-dimensional subspace.

Recently, an interesting work called discriminative locality alignment (DLA)[19,20] was proposed for DR. Generally, the main idea of DLA is described as two steps. First, discriminative information is imposed over patches, each of which is associated with one sample and its neighbors. Second, the alignment trick is used to align all of the part optimizations to the whole optimization. Although DLA shows impressive results, we find that the performance of DLA is sensitive to the variations of the model parameters.

Based on linear reconstruction, we propose locally reconstructive patch alignment (LRPA) for DR. By maximizing the local between-class reconstruction error and minimizing the local within-class reconstruction error simultaneously, LRPA preserves the local information for classification. More importantly, LRPA has natural connections with linear regression-based classification (LRC).[21] As we know, LRC assumes a sample belongs to the class with the minimum reconstruction error. Therefore, only the sample with the minimum within-class reconstruction error can be classified correctly. Since LRPA aims to find the projections by which the within-class reconstruction error is minimal, we can expect that LRC can be more effective in the LRPA's subspace.

The rest of the paper is organized as follows. Related works are reviewed in Sec. 2. In Sec. 3, LRPA is described in detail. In Sec. 4, the experiments are presented on the well-known databases to demonstrate the effectiveness and

0091-3286/2012/$25.00

robustness of the proposed method. Finally, conclusions are drawn in Sec. 5.

2 Discriminative Locality Alignment

DLA aims to extract discriminative information from patches. To achieve this goal, one patch is first built for each sample. Each patch includes a sample and its within-class nearest samples and its between-class nearest samples. Then an objective function is designed to preserve the local discriminative information of each patch. Finally, all the part optimizations are integrated together to form a global coordinate according to the alignment trick.[22] The projection matrix can be obtained by solving a standard Eigen decomposition problem.

2.1 *Part Optimization*

Suppose we have a set of samples $\mathbf{X} = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^{d\times N}$ from c different classes (where X = dataset in orginal space dataset in original space). For a given sample x_i (where $x_i = i$'th sample of $\mathbf{X}$), its k_1 (where k = number of neighbors) within-class nearest samples and k_2 between-class nearest samples are defined as $x_{i^1}, x_{i^2} \ldots, x_{i^{k_1}}$, and x_{i_1}, $x_{i_2} \ldots, x_{i_{k_2}}$, respectively. The local patch for the sample x_i is constructed by putting the sample x_i, its k_1 within-class nearest samples and k_2 between-class nearest samples together, i.e. (where $\mathbf{X}_i = i$'th patch),

$$\mathbf{X}_i = [x_i, x_{i^1}, x_{i^2}, \ldots, x_{i^{k_1}}, x_{i_1}, x_{i_2}, \ldots, x_{i_{k_2}}]. \tag{1}$$

For each patch, the corresponding output in the low-dimensional space is:

$$\mathbf{Y}_i = [y_i, y_{i^1}, y_{i^2}, \ldots, y_{i^{k_1}}, y_{i_1}, y_{i_2}, \ldots, y_{i_{k_2}}]. \tag{2}$$

In the low-dimensional space, we expect that distances between the given sample and its within-class samples are as small as possible, while distances between the given sample and its between-class samples are as large as possible. So we have

$$\begin{aligned} y_i &= \arg\ \min \sum_{j=1}^{k_1} \|y_i - y_{i^j}\|^2 \\ y_i &= \arg\ \min \sum_{p=1}^{k_2} \|y_i - y_{i_p}\|^2. \end{aligned} \tag{3}$$

Since the patch formed by the local neighborhood can be regarded approximately linear, we formulate the part discriminator by using the linear manipulation as follows:

$$y_i = \arg\ \min \left(\sum_{j=1}^{k_1} \|y_i - y_{i^j}\|^2 - \beta \sum_{p=1}^{k_2} \|y_i - y_{i_p}\|^2 \right), \tag{4}$$

where $\beta \in [0, 1]$ (where β = scaling factor) is the penalty parameter. Define the coefficients vector (where w_t = within-class reconstruction weight vector)

$$w_i = \left[\overbrace{1, \ldots, 1}^{k_1}, \overbrace{-\beta, \ldots, -\beta}^{k_2} \right]^T \tag{5}$$

then Eq. (4) reduces to:

$$\begin{aligned} &\arg\ \min \left[\sum_{j=1}^{k_1} \|y_i - y_{i^j}\|^2 w_i(j) \right. \\ &\left. - \sum_{p=1}^{k_2} \|y_i - y_{i_p}\|^2 w_i(p + k_1) \right] \\ &= \arg\ \min \left[\sum_{j=1}^{k_1+k_2} \|y_{F_i(1)} - y_{F_i(j+1)}\|^2 w_i(j) \right] \\ &= \arg\ \min\ \mathrm{tr}(\mathbf{Y}_i L_i \mathbf{Y}_i^T), \end{aligned} \tag{6}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator,

$$L_i = \begin{bmatrix} \sum_{j=1}^{k_1+k_2} w_i(j) & -w_i^T \\ -w_i & \mathrm{diag}(w_i) \end{bmatrix} \tag{7}$$

and $F_i = \{i, i^1, \ldots i^{k_1}, \ldots, i_1, \ldots, i_{k_2}\}$ is the index set for the i'th patch.

2.2 *Whole Alignment*

After the part optimization step, we unify the optimizations together as a whole one by assuming that the coordinate for the i'th patch $\mathbf{Y}_i = [y_i, y_{i^1}, y_{i^2}, \ldots, y_{i^{k_1}}, y_{i_1}, y_{i_2}, \ldots, y_{i_{k_2}}]$ is selected from the global coordinate $\mathbf{Y} = [y_1, y_2, \ldots, y_N]$ (where $\mathbf{Y}$ = dimension-reduced dataset), such that

$$\mathbf{Y}_i = \mathbf{Y}\mathbf{S}_i, \tag{8}$$

where $\mathbf{S}_i \in \mathbb{R}^{N\times(k_1+k_2+1)}$ is the selection matrix and an entry is defined as (where $\mathbf{S}_i$ = selection matrix):

$$(\mathbf{S}_i)_{\mathrm{pq}} = \begin{cases} 1 & \text{if } p = F_i\{q\} \\ 0 & \text{else} \end{cases}. \tag{9}$$

Then Eq. (6) can be rewritten as

$$\arg \min_{\mathbf{Y}} \mathrm{tr}(\mathbf{Y}\mathbf{S}_i L_i \mathbf{S}_i^T \mathbf{Y}^T). \tag{10}$$

By summing over all the part optimizations described as Eq. (10), we can obtain the whole alignment as

$$\arg \min_{\mathbf{Y}} \sum_{i=1}^{N} \mathrm{tr}(\mathbf{Y}\mathbf{S}_i L_i \mathbf{S}_i^T \mathbf{Y}^T) = \arg \min_{\mathbf{Y}} \mathrm{tr}(\mathbf{Y} L \mathbf{Y}^T), \tag{11}$$

where $L = \sum_{i=1}^{N} \mathbf{S}_i L_i \mathbf{S}_i^T \in \mathbb{R}^{N\times N}$ is the alignment matrix. It is obtained based on an iterative procedure (where L_i = representation of part optimization):

$$L(F_i, F_i) \leftarrow L(F_i, F_i) + L_i \tag{12}$$

for $i = 1, 2, \ldots, N$ (where N = size of the dataset) with the initialization $L = 0$. Note that $L(F_i, F_i)$ is a submatrix

constructed by selecting certain rows and columns from L according to the index set F_i. Therefore, Eq. (12) only updates a submatrix of L determined by the index set F_i in each iteration.

To obtain the linear and orthogonal projection matrix $\mathbf{P}$ (where $\mathbf{P}$ = projection matrix), such as $\mathbf{Y} = \mathbf{P}^T\mathbf{X}$, Eq. (11) is deformed as follows:

$$\arg\min_{\mathbf{P}} \operatorname{tr}(\mathbf{P}^T\mathbf{X}L\mathbf{X}^T\mathbf{P}), \quad \text{s.t.} \quad \mathbf{P}^T\mathbf{P} = \mathbf{I}. \tag{13}$$

The transformation matrix $\mathbf{P}$ that minimizes the objective function is given by the minimum eigenvalue solution to the standard eigenvalue problem,

$$\mathbf{X}L\mathbf{X}^T\mathbf{P} = \lambda\mathbf{P}. \tag{14}$$

3 Locally Reconstructive Patch Alignment

3.1 *Motivation*

Let $\mathbf{X} = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^{d\times N}$ be a training set from c classes (where c = number of classes). Suppose $\mathbf{X}_i^w = [x_{i^1}, x_{i^2}, \ldots, x_{i^{k_1}}]$ (where $\mathbf{X}_i^w$ =within-class nearest neighbor set) is the within-class nearest neighbor set of x_i and $\mathbf{X}_i^b = [x_{i_1}, x_{i_2}, \ldots, x_{i_{k_2}}]$ (where $\mathbf{X}_i^b$ = between class nearest neighbor set) is the between-class nearest neighbor set of x_i. Thus the patch of x_i can be represented as $\mathbf{X}_i = [x_i, \mathbf{X}_i^w, \mathbf{X}_i^b]$. In the algorithm of DLA, the distance from a sample x_i to its nearest neighbor x_j is defined as the Euclidean distance, i.e. (where d = dimension of x_i),

$$d(x_i, x_j) = \|x_i - x_j\|^2. \tag{15}$$

Then the total distance from x_i to its within-class nearest neighbors can be regarded as the distance from x_i to its own class

$$d_i^w = \sum_{j=1}^{k_1} \|x_i - x_{i^j}\|^2. \tag{16}$$

Recent studies[21,23,24] have shown that, as a distance measurement, the linear reconstruction error is more robust than the Euclidean distance. Instead of using the Euclidean distance, we employ the reconstruction error to investigate the distance from a sample to its nearest neighbors. Thus the distance from x_i to its within-class neighbors and can be defined as follows:

$$d_i^w = \left\| x_i - \sum_{j=1}^{k_1} x_{i^j} w_i^j \right\|_2, \tag{17}$$

where w_i^j is the within-class reconstruction weight.

We can rewrite Eq. (17) using the matrix form:

$$d_i^w = \left\| x_i - \sum_{j=1}^{k_1} x_{i^j} w_i^j \right\|_2 = \|x_i - \mathbf{X}_i^w w_i\|_2, \tag{18}$$

where $w_i = [w_i^1, w_i^2, \ldots, w_i^{k_1}]^T$ is the within-class reconstruction weight vector.

The optimal within-class reconstruction weight vector can be obtained by least-squares estimation (LSE):[25]

$$w_i = [(\mathbf{X}_i^w)^T\mathbf{X}_i^w]^{-1}(\mathbf{X}_i^w)^T x_i. \tag{19}$$

From the geometrical view, minimizing the within-class reconstruction error in Eq. (17) is finding a point $\hat{x}$, which is closest to x_i in the within-class nearest neighbor's space (the space spanned by the within-class nearest neighbors). Actually, $\hat{x}$ is the projection of x_i onto the within-class nearest neighbor's space. And the within-class reconstruction error d_i^w is actually the distance from x_i to the within-class nearest neighbor's space.

Similarly, we can define the distance from x_i to its between-class neighbors

$$d_i^b = \left\| x_i - \sum_{p=1}^{k_2} x_{i^j} \tilde{w}_i^p \right\|_2 = \|x_i - \mathbf{X}_i^b \tilde{w}_i\|_2, \tag{20}$$

where $\tilde{w}_i = [\tilde{w}_i^1, \tilde{w}_i^2, \ldots \tilde{w}_i^{k_2}]^T$ (where $\tilde{w}_i$ = between-class reconstruction weight vector) is the between-class reconstruction weight vector, which can be obtained by Eq. (21)

$$\tilde{w}_i = [(\mathbf{X}_i^b)^T\mathbf{X}_i^b]^{-1}(\mathbf{X}_i^b)^T x_i. \tag{21}$$

The between-class reconstruction error d_i^b is actually the distance from x_i to the between-class nearest neighbor's space.

3.2 *Part Optimization*

Our goal is to find a low-dimensional linear embedding of the data by virtue of the linear transformation

$$y_i = \mathbf{P}^T x_i, \quad \text{where } \mathbf{P} = \{\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_2, \ldots \boldsymbol{\varphi}_m\}. \tag{22}$$

In the low-dimensional subspace, we expect that distances from samples to their corresponding within-class nearest neighbor's space are as small as possible, while distances from samples to their corresponding between-class nearest neighbor's space are as large as possible.

To achieve this goal, for each patch in the low-dimensional subspace, the distance from Y_i to the within-class nearest neighbor's space should be as small as possible, so we have:

$$\arg\min_{y_i} \left\| y_i - \sum_{j=1}^{k_1} y_{i^j} w_i^j \right\|_2. \tag{23}$$

Meanwhile, the distance from y_i to the between-class nearest neighbor's space should be as large as possible, so we have:

$$\arg\max_{y_i} \left\| y_i - \sum_{p=1}^{k_2} y_{i_p} \tilde{w}_i^p \right\|_2. \tag{24}$$

We can unify the objective function in Eqs. (23) and (24) as follows:

$$\arg\min_{y_i}\left(\left\|y_i-\sum_{j=1}^{k_1}y_{i^j}w_i^j\right\|_2-\beta\left\|y_i-\sum_{p=1}^{k_2}y_{i_p}\tilde{w}_i^p\right\|_2\right), \quad (25)$$

where $\beta\in[0,1]$ is the scale parameter to unify the different measures of the within-class distance and the between-class distance.

Assume $\mathbf{Y}=[y_i,y_{i^1},y_{i^2},\ldots,y_{i^{k_1}},y_{i_1},y_{i_2},\ldots,y_{i_{k_2}}]$ is the patch of x_i in the low-dimensional subspace. Let $\tilde{w}_i=[\tilde{w}_i^1,\tilde{w}_i^2,\ldots\tilde{w}_i^{k_1}]^T$ and $\tilde{w}_i=[\tilde{w}_i^1,\tilde{w}_i^2,\ldots\tilde{w}_i^{k_2}]^T$ be the within-class reconstruction weight vector and the between-class reconstruction weight vector. The objective function in Eq. (25) can be rewritten as:

$$\begin{aligned}&\left\|y_i-\sum_{j=1}^{k_1}y_{i^j}w_i^j\right\|_2-\beta\left\|y_i-\sum_{p=1}^{k_2}y_{i^j}\tilde{w}_i^p\right\|_2\\&=\|\mathbf{Y}_ie-\mathbf{Y}_i\omega_i\|_2-\beta\|\mathbf{Y}_ie-\mathbf{Y}_i\tilde{\omega}_i\|_2\\&=\mathrm{tr}((\mathbf{Y}_ie-\mathbf{Y}_i\omega_i)(\mathbf{Y}_ie-\mathbf{Y}_i\omega_i)^T\\&\quad-\beta(\mathbf{Y}_ie-\mathbf{Y}_i\tilde{\omega}_i)(\mathbf{Y}_ie-\mathbf{Y}_i\tilde{\omega}_i)^T)\\&=\mathrm{tr}(\mathbf{Y}_i[(e-\omega_i)(e-\omega_i)^T-\beta(e-\tilde{\omega}_i)(e-\tilde{\omega}_i)^T]\mathbf{Y}_i^T)\\&=\mathrm{tr}(\mathbf{Y}_iL_i\mathbf{Y}_i^T),\end{aligned} \quad (26)$$

where

$$L_i=(e-\omega_i)(e-\omega_i)^T-\beta(e-\tilde{\omega}_i)(e-\tilde{\omega}_i)^T \quad (27)$$

$$e=\left[1,\overbrace{0,\ldots0}^{k_1+k_2}\right]^T,\qquad \omega_i=\left[0,w_i,\overbrace{0,\ldots,0}^{k_2}\right]^T,\qquad \text{and}$$

$$\tilde{\omega}_i=\left[\overbrace{0,\ldots,0}^{k_1+1},\tilde{w}_i,\right]^T.$$

It is easy to understand that the vectors e, ω_i and $\tilde{\omega}_i$ encode the contribution of each column of $\mathbf{Y}_i$ during the reconstruction step.

3.3 *Whole Alignment*

In the whole alignment procedure, LRPA does the same steps as DLA. Finally, the optimal solution of LRPA can be obtained by solving the Eigenvectors of Eq. (28) corresponding with the smallest eigenvalues.

$$\mathbf{X}L\mathbf{X}^T\mathbf{P}=\lambda\mathbf{P}, \quad (28)$$

where

$$L=\sum_{i=1}^{N}\mathbf{S}_iL_i\mathbf{S}_i^T\in\mathbb{R}^{N\times N} \quad (29)$$

$$(\mathbf{S}_i)_{\mathrm{pq}}=\begin{cases}1 & \text{if } p=F_i\{q\}\\0 & \text{else}\end{cases} \quad (30)$$

and $F_i=\{i,i^1,\ldots i^{k_1},\ldots,i_1,\ldots,i_{k_2}\}$ is the index set for the i'th patch.

3.4 *Algorithm of LRPA*

The procedure of LRPA is summarized in Table 1.

Table 1 The algorithm of LRPA.

Input: Sample matrix $\mathbf{X}=[x_1,x_2\ldots,x_N]$, the scale parameter β, the number of the within-class nearest neighbors k_1 and the number of the between-class nearest neighbors k_2.

Output: Transform matrix $\mathbf{P}_{\mathrm{LRPA}}$

1. Project the training samples into a PCA subspace spanned by its leading eigenvectors: $\hat{\mathbf{X}}=[\hat{x_1},\hat{x_2},\ldots,\hat{x_N}]=\mathbf{P}_{\mathrm{PCA}}^T\mathbf{X}$

2. For each sample $\hat{x}_i$, find its k_1 within-class nearest neighbors $\mathbf{X}_i^w$ and k_2 between-class nearest neighbors $\mathbf{X}_i^b$. Then construct the patch $\mathbf{X}_i=[\hat{x}_i,\mathbf{X}_i^w,\mathbf{X}_i^b]$

3. Compute the within-class reconstruction weight vector w_i by Eq. (19) and between-class reconstruction weight vector $\tilde{w}_i$ by Eq. (21).

4. Compute L_i by Eq. (27) and iteratively calculate L by Eq. (29).

5. Solve the generalized eigenvectors of $\hat{\mathbf{X}}L\hat{\mathbf{X}}^T\mathbf{P}=\lambda\mathbf{P}$ and construct the projections $\mathbf{P}=\{\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\ldots,\boldsymbol{\varphi}_m\}$ corresponding to the m smallest eigenvalues.

6. Output $\mathbf{P}_{\mathrm{LRPA}}=\mathbf{P}_{\mathrm{PCA}}\mathbf{P}$

Table 2 Details of the four databases.

Database	Size	Number of classes	Number of samples per class	Number of training sample per class
YALE-B	32×32	38	64	10
AR	50×40	120	26	5
FKP	55×110	100	12	5
Palmprint	64×64	100	6	3

4 Experiments

To evaluate the performance of the proposed method, we compare it with four DR methods (KPCA,[26] LDA, NPE, and DLA) over two classifiers, i.e., nearest neighbor classifier (NNC)[27] and LRC, on four well-known databases. The details of the databases are summarized in Table 2. For efficiency, on the four databases, PCA is first applied to reduce the dimensionality. Then the experiments are performed on the 150-dimensional PCA subspace. On the YALE-B, AR, and finger knuckle print (FKP) databases, all the experiments are repeated 50 times independently. On the PolyU palm print database, we only compare the palm print images of the two sessions once.

4.1 *Face Recognition*

4.1.1 *Experiments on the YALE-B database*

The YALE-B database[28] consists of 2414 frontal face images of 38 subjects under various lighting conditions. The database was divided in five subsets: subset 1 consisting of

Fig. 1 Sample images of one person from the YALE-B face database.

Table 3 The average maximal recognition rates on the YALE-B database.

	KPCA	LDA	NPE	DLA	LRPA
NNC	43.1 ± 1.0	80.0 ± 1.4	79.5 ± 2.1	78.3 ± 1.4	$\mathbf{80.1 \pm 1.4}$
Dimension	150	25	90	40	70
LRC	81.1 ± 1.1	84.1 ± 1.1	87.8 ± 1.5	89.4 ± 0.9	$\mathbf{92.8 \pm 0.7}$
Dimension	150	37	150	120	80

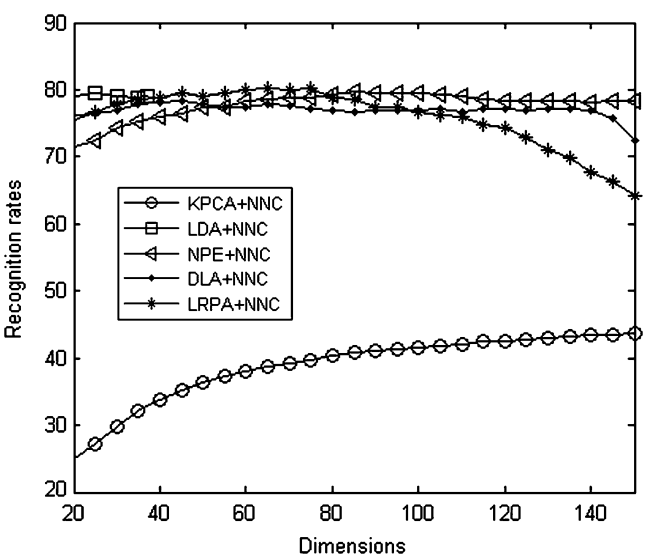


Fig. 2 Recognition rates curves of five methods plus NNC on the YALE-B database.

266 images (seven images per subject) under nominal lighting conditions was used as the gallery. Subsets 2 and 3, each consisting of 12 images per subject, characterize slight-to-moderate luminance variations, while subset 4 (14 images per person) and subset 5 (19 images per person) depict severe light variations. The images are also grayscale and normalized to a resolution of 32×32 pixels.

Sample images of one person from the YALE-B face database are shown in Fig. 1. The maximal recognitions and the standard deviations across 50 tests are listed in Table 3. And the recognition rates versus the dimensions with two classifications are illustrated in Figs. 2 and 3.

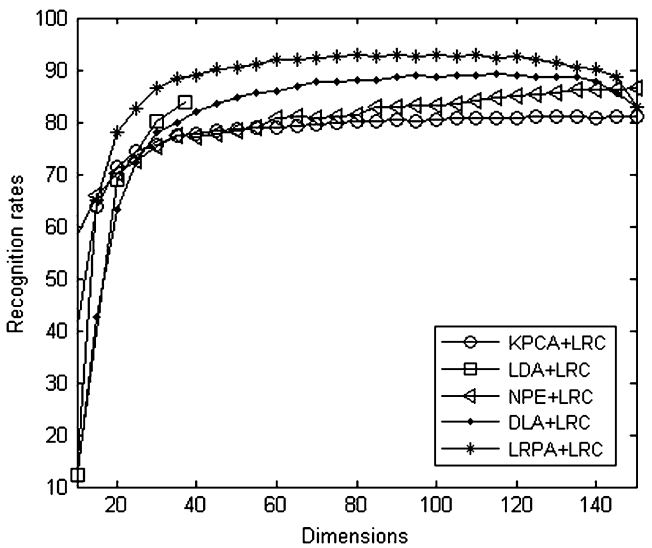


Fig. 3 Recognition rates curves of five methods plus LRC on the YALE-B database.

4.1.2 *Experiments on the AR database*

The AR face database[29] 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 most persons were taken in two sessions (separated by two weeks). Each section contains 13 color images and 120 individuals (65 men and 55 women) participated in both sessions. The images of these 120 individuals were selected and used in our experiment. Only the full facial images were considered here (no attempt was made to handle occluded face recognition in each session). We manually cropped the face portion of the image and then normalized it to 50×40 pixels.

Sample images of one person from the AR face database are shown in Fig. 4. The maximal recognition rates and the standard deviations across 50 tests are listed in Table 4. And the recognition rates versus the dimensions with two classifications are shown in Figs. 5 and 6.

4.2 *Finger Knuckle Print Recognition*

In PolyU FKP database,[30–32] FKP images were collected from 165 volunteers, including 125 males and 40 females. Among them, 143 subjects were 20 to 30 years old, and the others were 30 to 50 years old. The samples were 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. Therefore, 48 images from four

Fig. 4 Sample images of one person from the AR face database.

Table 4 The average maximal recognition rates on the AR database.

	KPCA	LDA	NPE	DLA	LRPA
NNC	57.6 ± 1.1	88.6 ± 1.0	86.9 ± 0.8	89.4 ± 0.8	$\mathbf{90.1 \pm 0.9}$
Dimension	150	110	120	110	70
LRC	71.1 ± 1.3	90.4 ± 1.1	92.2 ± 0.7	92.8 ± 0.8	$\mathbf{93.3 \pm 0.8}$
Dimension	150	110	120	120	80

fingers were collected from each subject. In total, the database contains 7920 images from 660 different fingers. The average time interval between the first and the second sessions was about 25 days. The maximum and minimum time intervals were 96 days and 14 days, respectively. All the samples in the database are histogram equalized and resized to 55×110.

Sample images of one person from the FKP database are shown in Fig. 7. The maximal recognition rates and the standard deviations across 50 tests are listed in Table 5. And the recognition rates versus the dimensions with two classifications are shown in Figs. 8 and 9.

4.3 *Palm Print Recognition*

The PolyU palm print database[33] contains 600 gray-scale 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.

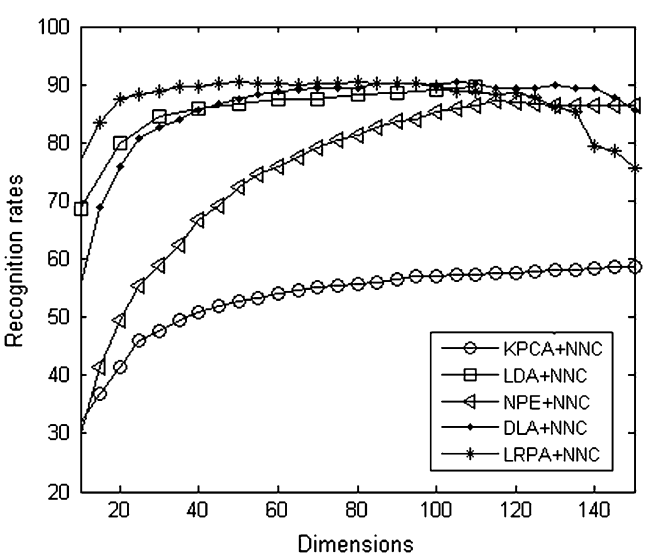


Fig. 5 Recognition rates curves of five methods plus NNC on the AR database.

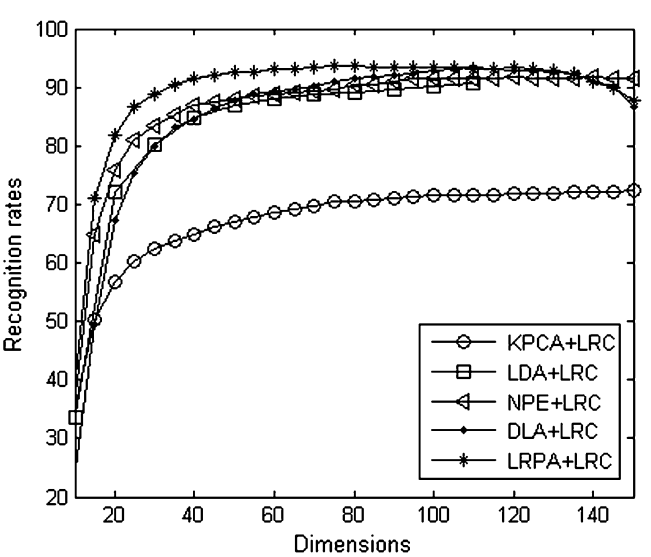


Fig. 6 Recognition rates curves of five methods plus LRC on the AR database.

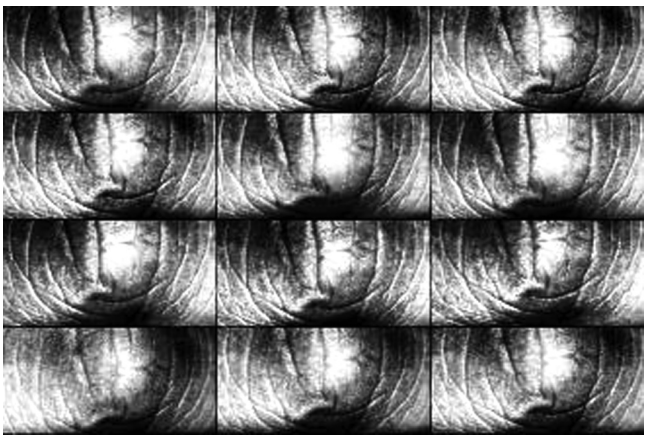


Fig. 7 Sample images of the right index finger from one individual.

The average interval between the first and the second sessions is two months. The images were resized to 64×64 pixels and preprocessed using histogram equalization. Figure 10 shows some sample images of the two sessions.

According to the protocol of this database, the palm print images are divided into two groups: one group is made up of three 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. The recognition rates are listed in Table 6. The recognition rates versus the dimensions with two classifications are shown in Figs. 11 and 12.

4.4 *Parameter Sensitivity Analysis*

In the model of DLA and LRPA, there are three parameters: β (the scale parameter), k_1 (the number of the within-class nearest neighbors), and k_2 (the number of the between-class nearest neighbors). In this subsection, we will investigate the effects of the three model parameters on the recognition rates in the validation phase based on the YALE-B database over two classifications (NNC and LRC). The selected subspace dimension was fixed to 80. In this experiment, 20 images of each person are used for training. Therefore, $1 \leq k_1 \leq 19$, $1 \leq k_2 \leq 740$, and $0 \leq \beta \leq 1$.

To evaluate the effects of the three model parameters, we fix two model parameters each time. Then we vary the third parameter to obtain the recognition rates. First, by fixing k_1 and k_2 to arbitrary values, we can obtain the recognition rate curve with respect to β as shown in Fig. 13. We observe that

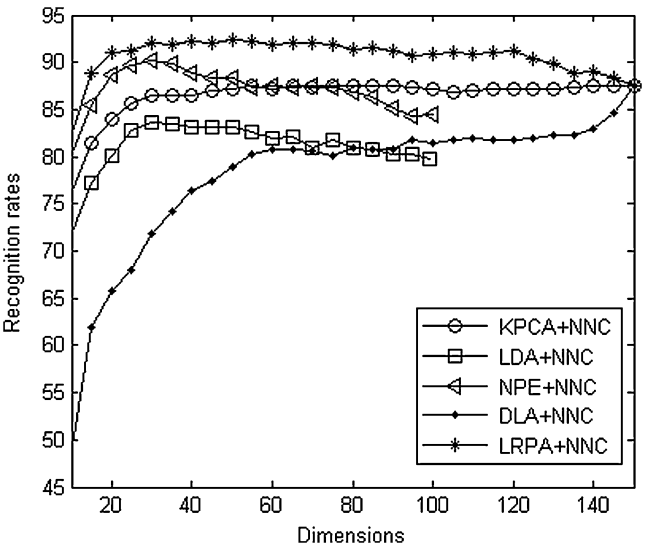


Fig. 8 Recognition rates curves of five methods plus NNC on the FKP database.

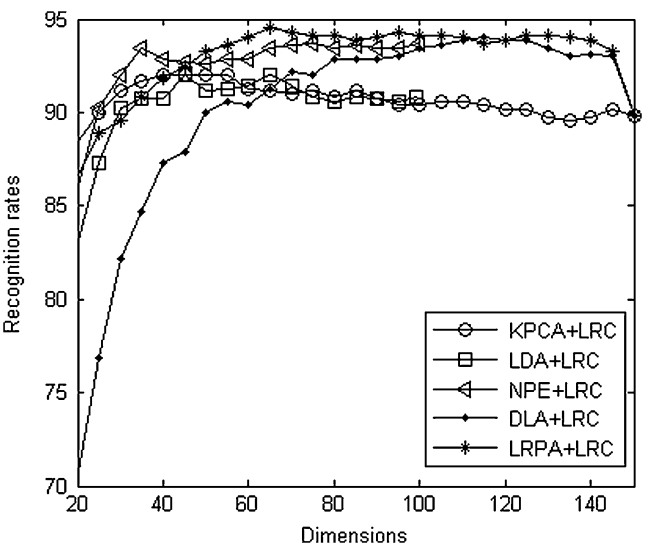


Fig. 9 Recognition rates curves of five methods plus LRC on the FKP database.

DLA achieves the highest recognition rates when $\beta = 0$, which means the between-class information may be unhelpful to the performance of DLA. Differently, LRPA achieves the highest recognition rates when $\beta = 2$, which means LRPA utilizes the between-class information effectively. Meanwhile, with the increment of β, we find that LRPA

Table 5 The average maximal recognition rates on the FKP database.

	KPCA	LDA	NPE	DLA	LRPA
NNC	86.8 ± 1.2	83.5 ± 1.6	89.5 ± 1.5	86.8 ± 1.5	$\mathbf{93.0 \pm 1.3}$
Dimension	150	30	30	150	50
LRC	91.6 ± 1.4	92.4 ± 1.3	93.8 ± 1.3	93.8 ± 1.3	$\mathbf{94.0 \pm 1.2}$
Dimension	50	30	120	120	65

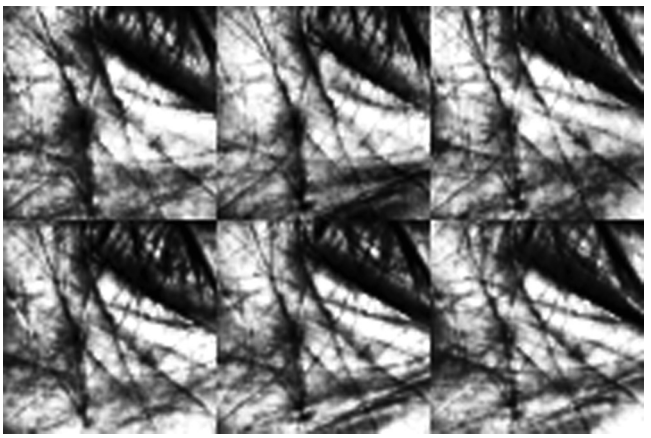


Fig. 10 Sample images of the two sessions in the PolyU palmprint database.

Table 6 The average maximal recognition rates on the PolyU palmprint database.

	KPCA	LDA	NPE	DLA	LRPA
NNC	88	**97.0**	96.7	96.7	96.3
Dimension	100	99	100	110	100
LRC	87.3	97.3	98.3	98.3	**99.3**
Dimension	150	99	100	120	90

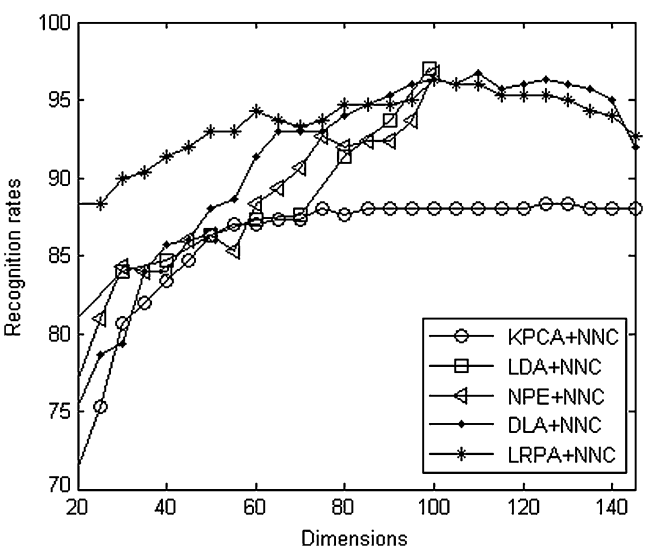


Fig. 11 Recognition rates curves of five methods plus NNC on the PolyU palmprint database.

is more stable than DLA, and the difference of their recognition rates is not significant.

Secondly, we fix $\beta = 0.2$ and set k_2 to arbitrary value. Then we compute the recognition rates against the variation of k_1. The recognition rate curves with respect to k_1 are drawn in Fig. 14. From Fig. 14, we find that the performance of LRPA plus NNC is not stable with the variation of k_1.

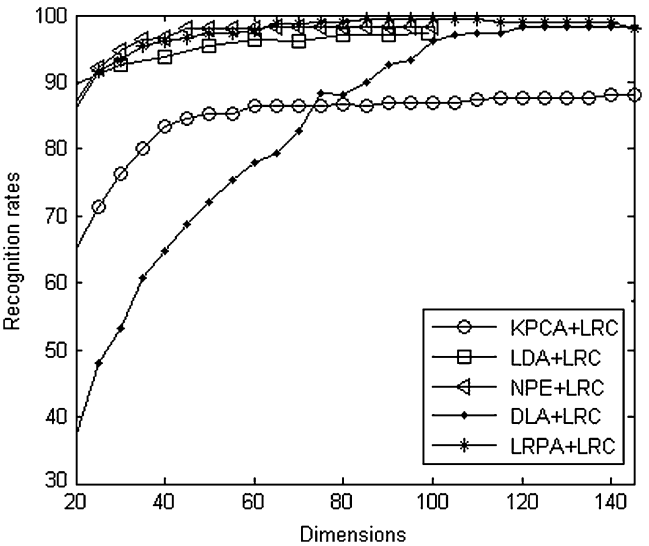


Fig. 12 Recognition rates curves of five methods plus LRC on the PolyU palmprint database.

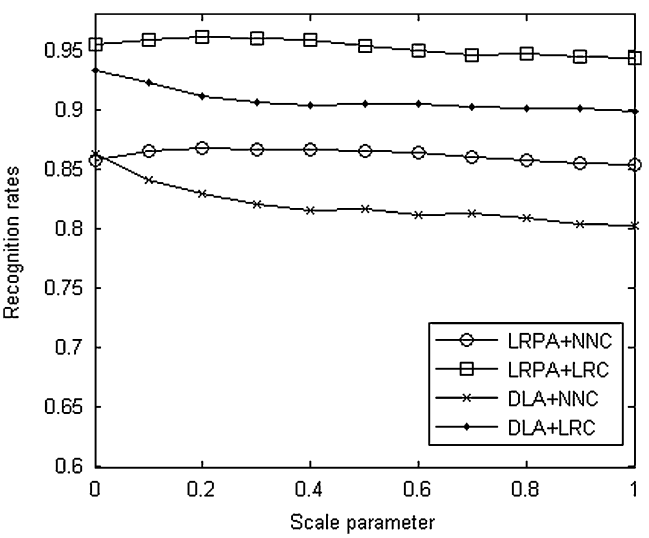


Fig. 13 Recognition rates versus scale parameter.

Third, we fix $\beta = 0.2$ and $k_1 = 5$. Then we show the recognition rate curves with respect to $k_1 = 5$ in Fig. 15.

4.5 *Discussion*

Based on the experimental results, we can draw the following conclusions:

(1) LRPA plus LRC outperforms other combinations. The experimental results in Tables 3–6 shows that LRPA plus LRC achieves higher recognition rates than those of other combinations.

(2) LRC is more suitable for LRPA than NNC. As can be seen from Fig. 14, the recognition rates of LRPA plus LRC are insensitive to the number of within-class nearest neighbors, whereas the recognition rates of LRPA plus NNC decrease severely with the growth of number of within-class nearest neighbors. Meanwhile, from Figs. 2 and 3, Figs. 5 and 6, Figs. 8 and 9,

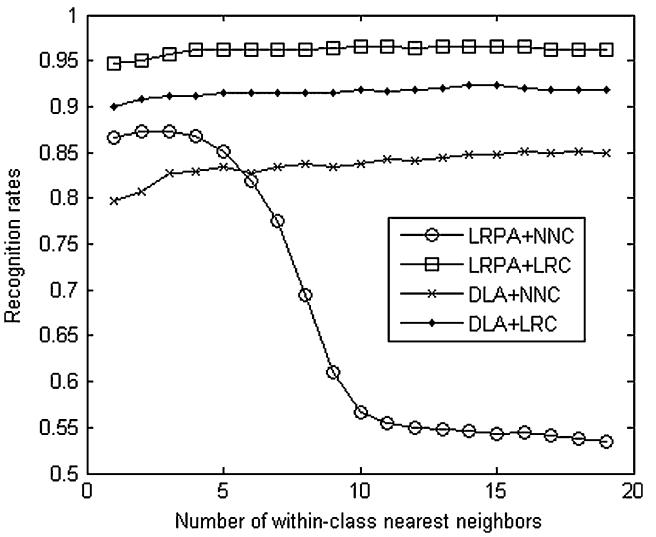


Fig. 14 Recognition rates versus number of within-class nearest neighbors.

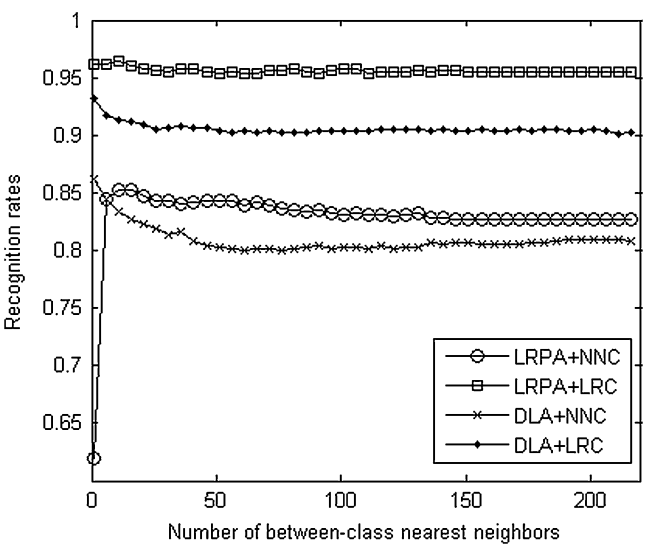


Fig. 15 Recognition rates versus number of between-class nearest neighbors.

and Figs. 11 and 12, we observe that the recognition rates of LRPA plus LRC are higher and more stable than those of LRPA plus NNC with the increment of dimensions. Technically, minimizing the reconstruction error of the within-class nearest neighbors cannot promise the given sample, and its nearest neighbor belong to the same class. Differently, LRC estimates the label of a sample based on the reconstruction error of each class. Therefore, minimizing the reconstruction error of within-class nearest neighbors is helpful to LRC.

(3) LRPA plus LRC is more robust to the variations of the model parameters. As can be seen from Figs. 13 through 15, the recognition rates of LRPA plus LRC vary in a very small range. Moreover, we notice that the between-class information may be not helpful to DLA. From Figs. 13 and 15, we can see that the recognition rates of DLA plus LRC or NN achieves the highest recognition rates when the scale parameter and the number of between-class nearest neighbors equals to zero. And the recognition rates of DLA decrease with the increment of scale parameter and the number of between-class nearest neighbors.

5 Conclusions

This paper introduces a new method called "locally reconstructive patch alignment for dimensionality reduction." Based on the local patch concept, LRPA minimizes the within-class nearest neighbor reconstruction error and maximizes the within-class nearest neighbor reconstruction error of the patches simultaneously. Compared with DLA, LRPA has the following advantages:

(1) LRPA has natural connections with LRC since both of LRPA and LRC are focus on reconstruction errors.

(2) LRPA plus LRC is more robust to the variations of parameters.

The experimental results on the YALE-B, AR, FKP, and palm print databases demonstrate the effectiveness and robustness of the proposed method.

Acknowledgments

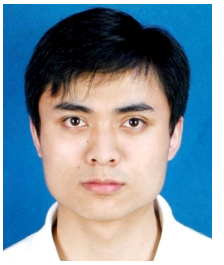
This work is partially supported by the National Science Foundation of China under Grant Nos. 90820306, 60873151, 60973098, and 61005008.

References

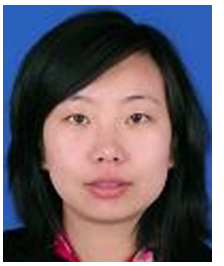
1. A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(1), 4–37 (2000).
2. I. T. Jolliffe, *Principal Component Analysis*, Springer, New York (1986).
3. P. N. Belhumeur, J. Hespanda, and D. Kiregeman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997).
4. L.-F. Chen et al., "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recogn.* **33**(10), 1713–1726 (2000).
5. D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(8), 831–836 (1996).
6. J. Yang and J. Yang, "Why can LDA be performed in PCA transformed space?," *Pattern Recogn.* **36**(2), 563–566 (2003).
7. J. Ye et al., "An optimization criterion for generalized discriminant analysis on undersampled problems," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(8), 982–994 (2004).
8. H. Cevikalp et al., "Discriminative common vectors for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(1) (2005).
9. H. Yu and J. Yang, "A direct LDA algorithm for high dimensional data-with application to face recognition," *Pattern Recogn.* **34**(10), 2067–2070 (2001).
10. Z. Jin et al., "Face recognition based on the uncorrelated discrimination transformation," *Pattern Recogn.* **34**(7), 1405–1416 (2001).
11. S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science* **290**(5500), 2323–2326 (2000).
12. J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science* **290**(5500), 2319–2323 (2000).
13. X. He et al., "Neighborhood preserving embedding," in *Proc. of the 10th IEEE International Conf. on Computer Vision*, pp. 1208–1213, IEEE, Beijing (2005a).
14. X. He et al., "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(3), 328–340 (2005).
15. S. Yan et al., "Graph embedding and extension: a general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(1), 40–51 (2007).
16. H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *IEEE Conf. on Computer Vision and Pattern Recognition 2005 (CVPR 2005)*, pp. 846–853, IEEE, Beijing (2005).

17. J. Yang et al., "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).
18. R. Wang et al., "Maximal linear embedding for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(9), 1776–1792 (2011).
19. T. Zhang et al., "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.* **21**(9), 1299–1313 (2009).
20. T. Zhang, D. Tao, and J. Yang, "Discriminative locality alignment," in *Proc. 10th European Conf. on Computer Vision*, pp. 725–738, Springer, Marseille (2008).
21. I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(11), 2106–2112 (2010).
22. D. Zhao, Z. Lin, and X. Tang, "Laplacian PCA and its applications," in *Proc. of IEEE Int. Conf. on Computer Vision*, pp. 1–8, IEEE, Rio de Janeiro (2007).
23. J. Wright et al., "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009).
24. L. Zhang, M. Yang, and X. C. Feng, "Sparse representation or collaborative representation: which helps face recognition?," in *ICCV*, IEEE, Barcelona (2011).
25. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning; Data Mining, Inference and Prediction*, Springer, New York (2001).
26. B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.* **10**(5), 1299–1319 (1998).
27. T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inform. Theor.* **13**(1), 21–27 (1967).
28. K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(5), 684–698 (2005).
29. A. M. Martinez and R. Benavente, "The AR face database," CVC technical report, no. 24, (June 1998).
30. L. Zhang et al., "Online finger-knuckle-print verification for personal authentication," *Pattern Recogn.* **43**(7), 2560–2571 (2010).
31. L. Zhang, L. Zhang, and D. Zhang, "Finger-knuckle-print: a new biometric identifier," in *Proc. of the IEEE International Conf. on Image Processing*, IEEE, Cairo (2009).
32. L. Zhang, The FKP database: http://www.comp.polyu.edu.hk/~biometrics/FKP.htm.
33. L. Zhang, The palmprint database: http://www4.comp.polyu.edu.hk/~biometrics/.

Yi Chen received a BS in the School of Computer Science and Technology, Nanjing University of Science and Technology (NUST), China, in 2003. Now he is pursuing a PhD in pattern recognition and intelligent systems from Nanjing University of Science and Technology. His current research interests include pattern recognition, computer vision, and image processing. E-mail: cy1story@qq.com



Jun Yin received a BS in mathematics and a PhD in pattern recognition and intelligence systems from Nanjing University of Science and Technology (NUST), Nanjing, China, in 2006 and 2011, respectively. Now he is a lecturer in the College of Information Engineering at Shanghai Maritime University (SMU), Shanghai, China. He is the author of more than 10 scientific papers in pattern recognition. His current interest includes pattern recognition, machine learning, and face recognition.



Jie Zhu received a BS in the School of Computer Science and Technology, Nanjing University of Science and Technology (NUST), China, in 2004. Now she is pursuing a PhD in pattern recognition and intelligent systems from Nanjing University of Science and Technology. Her current research interests include pattern recognition, computer vision, and image processing.

Zhong Jin received a BS in mathematics, an MS in applied mathematics and the PhD degree in pattern recognition and intelligence system from 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. He also visited the Centre de Visioper 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: mailto:zhongjin@mail.njust.edu.cn zhongjin@mail.njust.edu.cn