Computing and Informatics, Vol. 31, 2012, 1465–1479

SUPERVISED KERNEL LOCALLY PRINCIPLE COMPONENT ANALYSIS FOR FACE RECOGNITION

Yongfeng Qi

College of Computer Science and Engineering
Northwest Normal University
967 Anning East Road, 73 007 0 Lanzhou, P.R. China
e-mail: yongfeng_qi@163.com

Jiashu Zhang

Sichuan Province Key Lab of Signal and Information Processing
Southwest Jiaotong University
111 Section 1 of the Second Ring Road North, 61 0031 Chengdu, P.R. China
e-mail: jszhang@home.swjtu.edu.cn

Communicated by R. Tadeusiewicz

Abstract. In this paper, a novel algorithm for feature extraction, named supervised kernel locally principle component analysis (SKLPCA), is proposed. The SKLPCA is a non-linear and supervised subspace learning method, which maps the data into a potentially much higher dimension feature space by kernel trick and preserves the geometric structure of data according to prior class-label information. SKLPCA can discover the nonlinear structure of face images and enhance local within-class relations. Experimental results on ORL, Yale, CAS-PEAL and CMU PIE databases demonstrate that SKLPCA outperforms EigenFaces, LPCA and KPCA.

Keywords: Kernel trick, within-class geometric structure, principal component analysis, face recognition

Mathematics Subject Classification 2010: 68T10, 68U10, 68T45

1 INTRODUCTION

For face recognition task, feature extraction is a crucial step which is used to reduce dimensionality of data and to enhance the discriminatory information. Principal component analysis (PCA) is a well-known feature extraction method and aims to find a low-dimension space that captures the directions of maximizing variance in the data. EigenFaces [1] is a famous face recognition method which used PCA to reduce the dimensionality of sample data. PCA is one of linear feature extraction methods. Linear feature extraction methods often are inadequate to describe complex nonlinear variations of face images because of illumination, pose, and facial expression changes. To address this problem, kernel based techniques are often used to discover the nonlinear structure of the face images [2, 3, 4, 5, 6, 7]. Kernel Principal Component Analysis (KPCA) [8], which combines kernel trick with PCA, is a nonlinear extension of PCA. The main idea of KPCA is to first map the input space into a high-dimension feature space using a nonlinear mapping, then the principal components are computed in a feature space. KPCA has already shown to provide better recognition performance than PCA for face recognition task [9].

PCA and KPCA seek to find the global structure information of data. Recent researches show that the local structure information plays important role in face recognition task [10, 11, 12, 13, 14, 15]. Motivated by the idea of LPP [16], Yang et al. [17] proposed locally principal component analysis (LPCA) technique which seeks to discover the local structure information of data by the nearest neighbors. However, LPCA may result in the overlap of different class samples in feature space, because the nearest neighbors may belong to different class due to the influence of lighting, expression, pose and viewpoint.

KPCA and LPCA are unsupervised learning algorithms which do not take label information into consideration. For recognition task, the prior class-label information is generally important. In this paper, we present a supervised nonlinear feature extraction method, named supervised kernel locally principal component analysis (SKLPCA). Firstly, nonlinear kernel mapping is used to map the data into a feature space. Then a linear transformation is obtained in feature space; the linear transformation preserves the within-class geometric structures of original data by embedding within-class neighbor graph. Thus, our method can not only present complex nonlinear variations of real face images but also enhance local within-class relations.

The rest of this paper is arranged as follows: LPCA is reviewed in Section 2. SKLPCA is presented in Section 3. Experiment results are reported in Section 4, and conclusions are given in Section 5.

2 LOCALLY PRINCIPAL COMPONENT ANALYSIS

Given a set of M training samples $A = \{x_1, x_2, \ldots, x_M\}$ in $\mathbf{R}^n$, LPCA aims to seek a transformation matrix $W = [w_1, w_2, \ldots, w_d]$ to map these samples to a set of points $\{y_1, y_2, \ldots, y_M\}$ in $\mathbf{R}^d (d \ll n)$ by maximizing the local covariance matrix:

$$\begin{aligned} S_L &= \frac{1}{2MK} \sum_{i=1}^{M} \sum_{j=1}^{M} H_{ij}(x_i - x_j)(x_i - x_j)^T \\ &= \frac{1}{MK} (\sum_{i=1}^{M} D_{ii} x_i x_i^T - \sum_{i=1}^{M} \sum_{j=1}^{M} H_{ij} x_i x_j^T \\ &= \frac{1}{MK} (XDX^T - XHX^T) \\ &= \frac{1}{MK} XLX^T \end{aligned} \tag{1}$$

where H_{ij} is an adjacency matrix which is defined as follows:

$$H_{ij} = \begin{cases} 1, & \text{if } x_j \text{ is among the } K\text{-nearest of } x_i \\ 0, & \text{otherwise.} \end{cases}$$

The optimal transformation matrix W is composed of the orthonormal eigenvectors $w_1, w_2, \ldots, w_d$ of S_L corresponding to the first d largest eigenvalues, i.e. $W = [w_1, w_2, \ldots, w_d]$.

3 SUPERVISED KERNEL LOCALITY PRINCIPAL COMPONENT ANALYSIS (SKLPCA)

3.1 Principle

The nonlinear mapping Φ is used to map the input data $x \in \mathbf{R}$ into a feature space F. For a given set of M training samples $x_1, x_2, \ldots, x_M$ in input space, the mapped data in feature F are $\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_M)$. The local covariance matrix in feature space is defined as follows:

$$\begin{aligned} \tilde{S}_L &= \frac{1}{2MK} \sum_{i=1}^{M} \sum_{j=1}^{M} H_{ij}(\Phi(x_i) - \Phi(x_j))(\Phi(x_i) - \Phi(x_j))^T \\ &= \frac{1}{MK} QLQ^T \end{aligned} \tag{2}$$

where $Q = [\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_M)]$. Different from LPCA, in our proposed method, the adjacency matrix H_{ij} is defined as follows:

$$H_{ij} = \begin{cases} e^{-\frac{\left\|x_i - x_j\right\|^2}{t}}, & \text{if } x_i \text{ and } x_j \text{ belong to the same class and } x_j \\ & \text{is among the } k\text{-nearest of } x_j \\ 0, & \text{otherwise.} \end{cases}$$

Due to using prior class label information, our proposed SKLPCA is a supervised learning algorithm. Comparing to unsupervised methods (such as LPCA), our algorithm can make efficient use of label information of samples. Therefore, our method can better describe the geometrical structure in the data.

SKLPCA aims to find a projection axis where variance of mapped vectors in feature space is maximum, i.e.

$$\begin{aligned} J(\tilde{w}) = &\ \tilde{w}^T \tilde{S}_L \tilde{w} \\ = &\ \tilde{w}^T \left[\frac{1}{2MK} \sum_{i=1}^{M} \sum_{j=1}^{M} H_{ij} (\Phi(x_i) - \Phi(x_j))(\Phi(x_i) - \Phi(x_j))^T \right] \tilde{w} \\ = &\ \frac{1}{2MK} \sum_{i=1}^{M} \sum_{j=1}^{M} H_{ij} \left[\tilde{w}^T (\Phi(x_i) - \Phi(x_j))(\Phi(x_i) - \Phi(x_j))^T \tilde{w} \right] \\ = &\ \frac{1}{2MK} \sum_{i=1}^{M} \sum_{j=1}^{M} H_{ij} (y_i - y_j)(y_i - y_j)^T \\ = &\ \frac{1}{2MK} \sum_{i=1}^{M} \sum_{j=1}^{M} H_{ij} (y_i - y_j)^2 \end{aligned} \tag{3}$$

where y_i and y_j are projected vectors of $\Phi(x_i)$ and $\Phi(x_j)$, respectively. With our choice of the adjacency matrix H_{ij}, if x_i and x_j are closer then H_{ij} is greater, which ensures that y_i and y_j have a greater contribution to $J(\tilde{(w)})$. In this sense, SKLPCA preserves the within-class geometric structures of original data.

According to the theory of producing kernel [18], any $\tilde{w}$ can be expressed by a line combination of $\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_M)$:

$$\tilde{w} = \sum_{i=1}^{M} z_i \Phi(x_i) = Qz \tag{4}$$

where$z = (z_1, z_2, \ldots, z_M)^T$. Combining with (4), $J(\tilde{(w)})$ can be written as:

$$\begin{aligned} J(\tilde{w}) &= \tilde{w}^T \tilde{S}_L \tilde{w} \\ &= \frac{1}{MK} (Qz)^T Q L Q^T (Qz) \\ &= \frac{1}{MK} z^T (Q^T Q) L (Q^T Q) z. \end{aligned} \tag{5}$$

Matrix $R = Q^T Q$ is an $M \times M$ Gram matrix whose elements are determined by Equation (6):

$$R_{ij} = \Phi(x_i)^T \Phi(x_j) = (\Phi(x_i) \bullet \Phi(x_j)) = k(x_i, xj) \tag{6}$$

Thus, Equation (5) can be rewritten as:

$$J(z) = \frac{1}{MK} z^T RLRz = z^T \tilde{C} z \tag{7}$$

where $\tilde{C} = \frac{1}{MK} RLR$. $z = [Z_1, Z_2, \ldots, z_d]$ can be obtained by solving the eigenvalue problem:

$$\lambda_i z_i = \tilde{C} z_i, \qquad i = 1, 2, \ldots, d \tag{8}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ is the first d largest eigenvalues of $\tilde{C}$ and z_i is the corresponding eigrnvectors.

3.2 Feature Extraction

For a given sample $x \in \mathbf{R}^n$, its feature vector can be obtained by projecting its mapped vector $\Phi(x)$ into transformation matrix $\tilde{W} = [\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_d]$:

$$\begin{aligned} y &= [\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_d]^T \Phi(x) \\ &= [\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_d]^T Q^T \Phi(x) \\ &= Z^T [k(x_1, x), k(x_2, x), \ldots, k(x_M, x)]. \end{aligned} \tag{9}$$

3.3 Algorithm

Based on the above descriptions, the algorithmic procedure of SKLPCA can be given as follows:

Step 1: Construct the adjacency matrix. Let H denote an adjacency matrix whose elements are defined as follows:

$$H_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & \text{if } x_i \text{ and } x_j \text{ belong to the same class} \\ & \text{and } x_j\ k\text{-nearest of } x_j \\ 0, & \text{otherwise.} \end{cases}$$

Step 2: Compute diagonal matrix D (diagonal elements $D_{ii} = \sum_{i=1}^{M} H_{ij}$) and Laplacian matrix $L = D - H$.

Step 3: Compute Gram matrix $R = Q^T Q$, where $Q = [\Phi(x_1), \Phi(X_2), \ldots, \Phi(x_M)]$, and matrix $\tilde{C} = \frac{1}{MK} RLR$.

Step 4: Obtain matrix $Z = [z_1, z_2, \ldots, z_d]$ by solving eigenvalue problem: $\lambda_i z_i = \tilde{C} z_i, \qquad i = 1, 2, \ldots, d$

Step 5: For an input sample x, compute its feature vector: $y = Z^T [k(x_1, x), k(x_2, x), \ldots, k(x_M, x)]$.

4 EXPERIMENT RESULTS

In this section, our algorithm is evaluated and compared with PCA, KPCA, LPCA, and FisherFaces [19] methods on four face databases: ORL, Yale, CAS-PEAL and CMU PIE. The ORL database is used to evaluate the performance under the conditions where both sample size and pose are varied. The Yale database is used to examine the performance when both facial expressions and lighting are varied. CMU PIE database is used to check the performance under conditions where lightings are varied. The CAS-PERL is used to value the performance under conditions where there are variations in facial expressions. In all the experiments, the nearest neighbor classifier (Euclidean distance) is used for classification purposes.

4.1 Experimental Results on ORL Face Database

The ORL database (`http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`) contains 400 images from 40 individuals and each has 10 different images. For some individuals, the images are varied in lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). The size of each image is 112×92 pixels, with 256 grey levels per pixel. Figure 1 shows samples of one person on ORL database.

Fig. 1. Sample images of one person on ORL database

Firstly, we test the effect of different kernel functions on the recognition performance. First five images per person are selected as training set and the remaining images for test. Here, polynomial kernel function and Gaussian kernel function are taken into consideration. Polynomial kernel function and Gaussian kernel function are given in Equations (10) and (11), respectively:

$$k(x, y) = (x \bullet y + 1)^d \tag{10}$$

$$k(x, y) = \exp\left(- \|x - y\|^2\right)/\sigma \tag{11}$$

where parameter σ is set as $t \times n$ (n is the dimension of image vector). In this experiment, parameters t and d vary form 0.1 to 10. Figures 2 and 3 show recognition rates of SKLPCA when kernel functions are polynomial kernel function and Gaussian kernel function, respectively.

We can see from Figure 2 that the recognition rates drop slowly with increasing parameter d of polynomial kernel function. We also find that the recognition rates are very poor when parameter t of Gaussian kernel function is smaller than 0.3

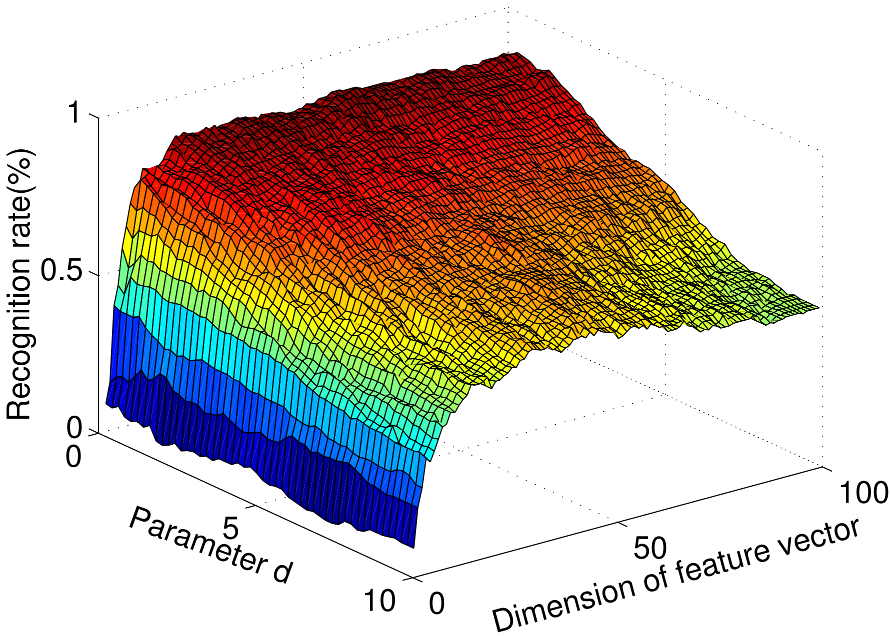


Fig. 2. Effect of kernel function on recognition performance of SKLPCA with polynomial kernel

from Figure 3, but the recognition rates are significantly improved when t is larger than 0.3.

We also compare SKLPCA with KPCA when kernel functions are Gaussian kernel and polynomial kernel, respectively. Table 1 presents the top recognition rates (%) of SKLPCA and KPCA with various parameters t or d; the values in parentheses denote the corresponding dimension of feature vector. We can see that the selection of kernel functions and different parameter values affect the recognition performances. Due to taking the prior class-label information into consideration, SKLPCA outperforms KPCA in term of recognition performance.

Secondly, we evaluate the recognition performance of SKLPCA and compare it with other methods: PCA, KPCA, LPCA, and FisherFaces when the number of training samples is varied. We construct five training sets by randomly selecting 2, 3, 4, 5 and 6 images of each person. The remaining images are used for test. The recognition procedure is repeated 10 times by randomly choosing different training and testing sets. Table 2 lists the top average recognition rates and standard deviation of five methods. The values in parentheses denote the corresponding dimension of feature vectors. Polynomial kernel function (parameter $d = 0.5$) is adopted for SKLPCA and KPCA. We can see from Table 2 that the nonlinear methods (KPCA

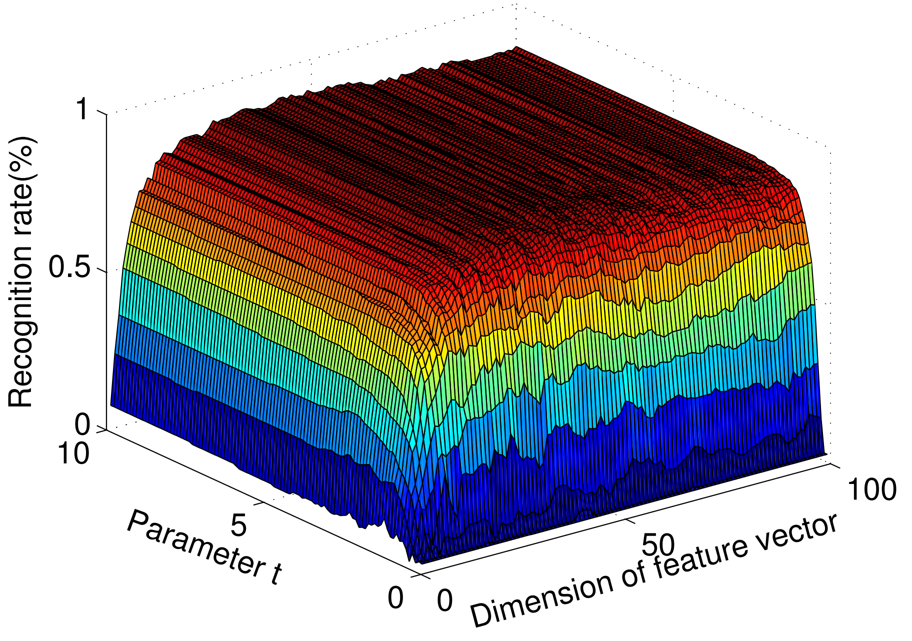


Fig. 3. Effect of kernel function on recognition performance of SKLPCA with Gaussian kernel

Methods	Polynomial Kernel				Gaussian Kernel			
	$d = 0.1$	$d = 0.5$	$d = 1.0$	$d = 1.5$	$t = 0.1$	$t = 0.5$	$t = 1.0$	$t = 1.5$
SKLPCA	91.00 (57)	91.00 (51)	91.50 (50)	92.50 (49)	45.00 (19)	90.50 (60)	92.50 (50)	93.00 (51)
KPCA	89.50 (92)	91.00 (75)	90.00 (73)	88.00 (89)	12.50 (40)	89.00 (66)	89.00 (61)	89.50 (75)

Table 1. Result comparison on comparison of SKLPCA and KPCA with different kernel function on ORL database

and SKLPCA) outperform linear methods (PCA, LPCA, FisherFacces); the main cause is the nonlinear methods can better model variations of poses which results in nonlinear distribution of face images. Moreover, our SKLPCA method also preserves within-class geometric structures of original data by embedding within-class nearest neighbors, which explains the reason why our method is better than KPCA.

Methods	Number of training samples				
	2	3	4	5	6
SKLPCA	81.88 ± 2.79 (38)	89.79 ± 2.69 (40)	93.50 ± 2.05 (46)	95.80 ± 1.17 (39)	97.37 ± 1.12 (55)
KPCA	79.00 ± 3.59 (42)	86.36 ± 1.66 (46)	92.00 ± 1.43 (55)	94.30 ± 1.35 (46)	95.00 ± 0.77 (41)
LPCA	79.50 ± 6.36 (35)	85.40 ± 2.01 (38)	90.11 ± 1.67 (41)	92.25 ± 1.31 (37)	94.25 ± 2.69 (38)
PCA	79.81 ± 1.26 (58)	87.07 ± 1.79 (56)	90.50 ± 1.60 (57)	93.40 ± 1.35 (59)	94.50 ± 1.28 (57)
FisherFaces	79.34 ± 2.26 (26)	87.75 ± 1.52 (39)	91.42 ± 1.91 (38)	93.30 ± 1.89 (39)	94.17 ± 1.78 (39)

Table 2. The top average recognition rates (%) and standard deviation of five methods under varying number of training samples on ORL database

4.2 Experimental Results on Yale Database

Yale database (`http://cvc.yale.edu/projects/yalefaces/yalefaces.html`) contains 165 images from 15 individuals, each has 11 images with varying facial expression or lighting. The size of each image is 320×243 pixels, with 256 grey levels per pixel. In our experiment, the images are manually cropped and resized to 64×64 pixels and no other preprocessing is conducted. We considered this database in order to evaluate the performance of algorithms under the condition where facial expression and lighting conditions are varied. Sample images of one person are given in Figure 4.

Fig. 4. Sample images of one person on Yale database

Firstly, we check the performance of the proposed SKLPCA under noise conditions. First five images per person are selected for training and the remaining images for test. All images in the test set are added with Gaussian noise (mean$\mu = 0$ and stand deviation $\sigma = 0.1$), salt & pepper noise (the noise density $\rho = 0.1$) and multiplicative noise, respectively. Thus, we have three test sample sets with various noises. The multiplicative noise is added to the image using the equation $y = x + n * x$, where x denotes the clean image and n is uniformly distributed random noise with mean 0 and variance V. Polynomial kernel function (parameter $d = 0.5$) is adopted for SKLPCA and KPCA. Table 3 lists the top recognition rates of five methods under three noise conditions and clean condition. The values in parentheses denote the corresponding dimension of feature vectors. We can find from Table 3 that SKLPCA outperforms other methods.

Methods	Without noise	Gaussian noise	salt & pepper noise	multiplicative noise
		$\sigma = 0.1$	$\rho = 0.1$	$V = 0.3$
SKLPCA	86.67 (36)	78.89 (45)	84.44 (41)	83.33 (36)
KPCA	81.11 (17)	70.00 (18)	75.56 (19)	80.00 (18)
LPCA	77.78 (29)	72.22 (36)	75.56 (47)	74.44 (31)
PCA	81.11 (17)	70.00 (16)	75.56 (17)	73.33 (18)
FisherFaces	83.33 (13)	76.67 (14)	81.11 (14)	81.11 (14)

Table 3. Comparison of recognition performance under noise conditions on Yale database

Secondly, we test the recognition performance of five methods under varying number of training samples. We construct five training sets by randomly selecting 2, 3, 4, 5 and 6 images of each person. The remaining images are used for test. As mentioned previously, we repeated each experiment 10 times. Table 4 shows the top average recognition accuracy and correspondence to standard deviation. Table 4 reveals that supervised learning algorithms (SKLPCA and FisherFaces) methods outperform unsupervised methods (PCA, LPCA and KPCA). We can see that our proposed SKLPCA method is comparable to FisherFaces under conditions which both lighting and facial expressions varied.

Methods	Number of training samples				
	2	3	4	5	6
SKLPCA	56.56 ± 2.76 (15)	74.83 ± 2.67 (29)	78.95 ± 1.93 (42)	83.89 ± 3.68 (55)	86.67 ± 4.26 (65)
KPCA	57.63 ± 2.96 (25)	59.58 ± 5.93 (38)	64.11 ± 2.90 (50)	65.56 ± 2.43 (66)	68.53 ± 4.8 (72)
LPCA	55.40 ± 5.54 (25)	61.33 ± 3.02 (39)	65.05 ± 2.56 (59)	66.11 ± 4.39 (83)	68.13 ± 4.55 (90)
PCA	56.44 ± 0.73 (30)	61.25 ± 1.87 (45)	64.00 ± 1.02 (58)	65.67 ± 1.21 (72)	67.87 ± 2.06 (82)
FisherFaces	57.93 ± 3.63 (13)	73.08 ± 3.14 (14)	80.57 ± 3.76 (14)	82.79 ± 3.41 (14)	85.93 ± 3.64 (14)

Table 4. Comparison of recognition performance under varying number of training samples on Yale database

4.3 Experiment Results on CUM PIE Face Database

The CMU pose, illumination and expression (PIE) database [20] contains 41 386 images of 68 individuals. The images were acquired under different pose and variable illumination conditions and with different facial expression. In this experiment, we choose only the frontal face images (C27) under varying illumination conditions (without expression variations); each person has 45 images. The images are cropped and resized to 64×64 pixels. Some sample images of one person are shown in Figure 5. We consider this database in order to evaluate the performance of algorithms under the conditions where lighting is varied.

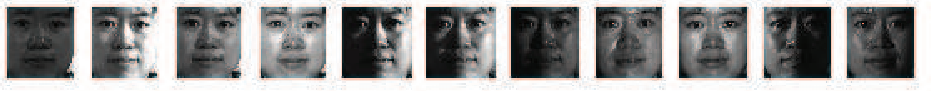


Fig. 5. some samples of one person on CMU PIE database

We randomly selected 25 images per person for training and the rest for test. The polynomial kernel function is selected for SKLPCA and KPCA, where $d = 0.5$. For each algorithm, the process is repeated 20 times and the average recognition rate curves with varying dimension of feature vectors are plotted in Figure 6. We can see from Figure 6 that supervised algorithms (SKLPCA and FisherFaces) outerperform unsupervised algorithms (KPCA, LPCA and PCA) under conditions where lightings are varied. We can also find that our proposed SKLPCA outperforms the FisherFaces method. In particular, the SKLPCA method achieves 99.1 percent correct face recognition accuracy when the dimension of feature vector is 88.

4.4 Experimental Result on CAS-PEAL Face Database

CAS-PEAL face database (`http://www.jdl.ac.cn/peal/index.html`) [21] contains 99 594 images of 1 040 individuals with varying pose, expression, accessories, and lighting. In this experiment, we choose the frontal faces with varying expressions and lightings to form a subset. This subset contains 910 images from 65 individuals; each has five various expression images and nine different lighting images. Samples of one individual are shown in Figure 7.

Firstly, we check the recognition performance of our proposed SKLPCA under conditions where facial expression are varied. We select the first two expression images per person for training and the remaining three expression images are used as test samples. Table 5 lists the top recognition accuracies of five methods with corresponding dimensions of feature vectors. The result indicates SKLPCA outperforms other methods under conditions where facial expressions are varied.

Secondly, we test the recognition performance of five algorithms on the full subset. We randomly select 3, 4, 5, 6, 7 samples to construct five training sets, and the remaining samples are used as test sets. The recognition procedure is repeated

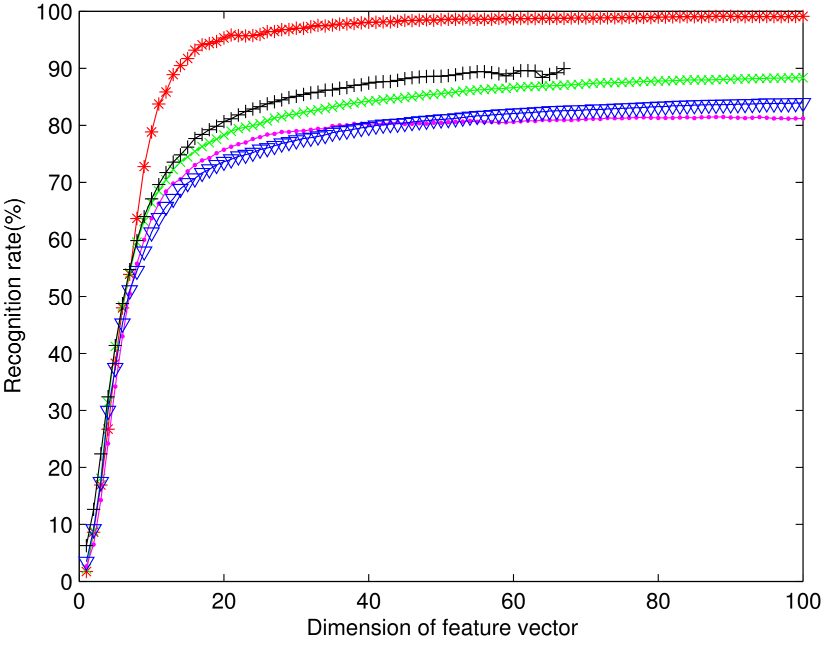


Fig. 6. Comparison of recognition rate of five methods under varying dimension on CMU PIE

Fig. 7. Samples of one person on CAS-PEAL: the first row shows images with varying expressions; the second row shows images with varying lighting conditions

Methods	SKLPCA	KPCA	LPCA	PCA	FisherFaces
The top recognition rate (%)	94.87 (65)	92.82 (59)	90.26 (36)	91.79 (45)	94.36 (51)

Table 5. The top recognition rates (%) of five methods under varying facial expressions on CAS-PEAL database

10 times by randomly choosing different training and testing sets. Table 6 gives the top average recognition rates (%) and standard deviation of five methods. The values in parentheses denote the corresponding dimension of feature vectors. We can find from Table 6 that the supervised methods (SKLPCA and FisherFaces) outperform unsupervised methods (KPCA, LPCA and PCA). And SKLPCA is comparable to FisherFaces in term of recognition performance.

Methods	Number of training samples				
	3	4	5	6	7
SKLPCA	70.32 ± 3.38 (94)	76.92 ± 1.33 (96)	82.57 ± 2.01 (96)	85.62 ± 1.53 (96)	89.63 ± 1.79 (97)
KPCA	45.96 ± 1.67 (96)	52.78 ± 1.32 (97)	58.36 ± 1.27 (96)	61.58 ± 2.51 (99)	65.00 ± 1.87 (100)
LPCA	44.34 ± 1.70 (96)	50.31 ± 1.78 (98)	57.67 ± 2.49 (90)	60.31 ± 3.46 (96)	64.88 ± 1.83 (95)
PCA	44.20 ± 1.24 (95)	51.50 ± 2.06 (96)	59.32 ± 1.89 (96)	61.23 ± 2.01 (95)	64.00 ± 2.37 (92)
FisherFaces	68.56 ± 2.23 (63)	76.98 ± 1.62 (64)	82.02 ± 1.30 (64)	85.46 ± 1.15 (63)	88.88 ± 2.01 (64)

Table 6. The top average recognition rates (%) and standard deviation of five methods under varying number of training samples on CAS-PEAL database

5 CONCLUSION

In this paper, a novel feature extraction method called SKLPCA is proposed. First, the face images are mapped into high dimension feature space by kernel trick. Then a linear transformation is obtained in feature space. The transformation preserves within-class geometric structures of the original data by embedding within-class neighbor graph. The proposed SKLPCA is a supervised nonlinear feature extraction method which can efficiently discover the nonlinear structure of face images. Experimental results on ORL, Yale, CMU PIE and CAS-PEAL databases show the effectiveness of our method.

5.1 Acknowledgements

The authors are grateful to the anonymous reviewers for their constructive comments and advices. The comments greatly improved the quality of the paper. This work is supported partly by the National Natural Science Foundation of China under Grant No. 61262056 and 51175443.

REFERENCES

[1] Turk, M.—Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience, Vol. 3, 1991, No. 1, pp. 71–86.

[2] Jing, X. Y.—Yao, Y. F.—Yang, J. Y.—Zhang, D.: A Novel Face Recognition Approach Based on Kernel Discriminative Common Vectors (KDCV) Feature Extraction and RBF Neural Network. Neurocomputing, Vol. 71, 2008, No. 13–15, pp. 3044–3048.

[3] Lu, J.—Plataniotis, K. N.–Venetsanopoulos, A. N.–Wang, J.: An Efficient Kernel Discriminant Analysis Method. Pattern Recognition, Vol. 38, 2005, No. 10, pp. 1788–1790.

[4] Chu, W. S.—Chen, J. C.—Lien, J. J.: Kernel Discriminant Transformation for Image Set-Based Face Recognition. Pattern Recognition, Vol. 44, 2011, No. 8, pp. 1567–1580.

[5] Hu, H.—Zhang, P.—Ma, Z.: Direct Kernel Neighborhood Discriminant Analysis for Face Recognition. Pattern Recognition Letters, Vol. 30, 2009, No. 10, pp. 902–907.

[6] Liu, Q.—Tang, X.—Lu, H.—Ma, S.: Face Recognition Using Kernel Scatter Difference-Based Discriminant Analysis. IEEE Transactions on Neural Networks, Vol. 17, 2006, No. 4, pp. 1081–1085.

[7] Baudat, G.—Anouar, F.: Generalized Discriminant Analysis Using a Kernel Approach. Neural Computation, Vol. 12, 2000, No. 10, pp. 2385–2404.

[8] Schölkopf, B.—Smola, A.—Müller, K.: Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation, Vol. 10, 1998, No. 5, pp. 1299–1319.

[9] Kim, K. I.—Jung, K.—Kim, H. J.: Face Recognition Using Kernel Principal Component Analysis. IEEE Signal Processing Letters, Vol. 9, 2002, No. 2, pp. 40–42.

[10] Zhu, L.—Zhu, S.: Face Recognition Based on Orthogonal Discriminant Locality Preserving Projections. Neurocomputing, Vol. 70, 2007, No. 7-9, pp. 1543–1546.

[11] Li, B.—Zheng, C. H.–Huang, D. S.: Locally Linear Discriminant Embedding: An Efficient Method for Face Recognition. Pattern Recognition, Vol. 41, 2008, No. 12, pp. 3813–3821.

[12] Yang, L.—Gong, W.—Gu, X.—Li, W.—Liang, Y.: Null Space Discriminant Locality Preserving Projections for Face Recognition. Neurocomputing, Vol. 71, 2008, pp. 3644–3649.

[13] Li, X.—Lin, S.—Yan, S.—Xu, D.: Discriminant Locally Linear Embedding with High-Order Tensor Data. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 38, 2008, No. 2, pp. 342–352.

[14] Xu, Y.—Song, F.—Feng, G.—Zhao, Y.: A Novel Local Preserving Projection Scheme for Use With Face Recognition. Expert Systems with Applications, Vol. 43, 2010, No. 6, pp. 1993–2002.

[15] Zhang, L.—Qiao, L.—Chen, S.: Graph-Optimized Locality Preserving Projections. Pattern Recognition, Vol. 43, 2010, No. 6, pp. 1993–2002.

[16] He, X.—Niyogi, P.: Locality Preserving Projections. Proceedings of the Conference on Advances in Neural Information Processing Systems 16, Vancouver 2003.

[17] Yang, J.—Zhang, D.—Yang, J.: Locally Principal Component Learning for Face Representation and Recognition. Neurocomputing, Vol. 69, 2006, No. 13–15, pp. 1697–1701.

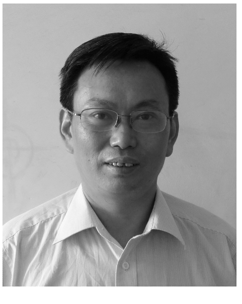
[18] Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York 1995.

[19] Belhumeur, P. N.—Hespanha, J. P.—Kreigman, D. J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 19, 1997, No. 7, pp. 711–720.

[20] Sim, T.—Baker, S.: The CMU Pose, Illumination, and Expression Database. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 25, 2003, No. 12, pp. 1615–1617.

[21] Gao, W.—Cao, B.—Shan, S.—Chen, X.—Zhou, D.—Zhang, X.—Zhao, D.: The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, Vol. 38, 2008, No. 1, pp. 149–161.

Yongfeng Qi received the Ph. D. degree from Southwest Jiaotong University in 2011. He is an Associate Professor at Northwest Normal University, China. His research interests include pattern recognition, computer vision and digital image processing.



Jiashu Zhang received the B. Sc. and Ph. D. degrees from the University of Electronic Science and Technology, China in 1987 and 2001, respectively. In 2001, he joined the School of Information Science and Technology, Southwest Jiaotong University, China. Currently he is is a Professor and the Director of the Key Laboratory of Signal and Information Processing of Sichuan Province. His research interests include digital signal processing, forensic information and data hiding, biometrics and security, and chaos theory with application to electronic engineering.