

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

A Kernel-based Collaborative Preserving Projection Based Face Recognition

Jing Zeng and Zhiwei Kang

School of Information Science and Engineering, Hunan University, Changsha, 410082, China

Abstract: Face recognition is an important issue in the field of pattern recognition, artificial intelligence and machine vision. The face image is not only sensitive to environment, illumination, pose and expression variation, but also has high-dimensional data disaster. In recently years, Sparse representation has been extensively studied in pattern recognition, which surprisingly pointed out that one target sample can be accurately represented as linear combination of very few measurement samples. Sparse Preserving Projection is a newly developed feature extraction method for reducing dimensions, which seeks a subspace where such sparse reconstructive relations among all training samples are preserved. However, Sparse Preserving Projection is time-consuming. In this study, a novel Kernel-based Collaborative Preserving Projection method is proposed for feature extraction. The method first uses a kernel-induced distance measure to determine k nearest neighbors of the target training sample from the remaining training samples and then the target training sample is reconstructed from its k nearest neighbors using collaborative representation. Finally the method seeks a low-dimensional subspace where the local collaborative reconstructive relations among all the training samples can be preserved. Experimental results on three benchmark databases show that the method can achieve a better classification result than Sparse Preserving Projection and reduce computation complexity.

Key words: Sparse representation, collaborative representation, spare preserving projection, face recognition

INTRODUCTION

Face Recognition (FR) involves many fields, mainly in image procession, pattern recognition, computer vision, artificial intelligence and so on. How to effectively extract and describe the facial feature is the key of face recognition research. As a result of the high dimensions in facial images, the direct representation of face features is very difficult. Dimension reduction plays an extensively important role in high-dimensional data analysis and studies. Classical dimensionality reduction methods including Principle Component Analysis (PCA) (Belhumeur *et al.*, 1997), Linear Discriminate Analysis (LDA) (Jing *et al.*, 2004) and Locality Preserving Projection (LPP) (He *et al.*, 2005) are successfully applied in pattern recognition. PCA projects the original high-dimensional data into a low-dimensional subspace where the data variety is maximized. PCA is an unsupervised algorithm. Contrarily, LDA is a supervised dimensionality reduction technique. It seeks a low-dimensional subspace where the data points of different classes are far from each other while data points from the same classes are clustered together. As a linear version of Locally Linear Embedding (LLE) (Roweis and Saul, 2000), LPP aims to find an embedded subspace which can preserve local neighborhood relations among samples. Besides, quite a lot of nonlinear

feature extraction methods related to PCA are proposed, such as Kernel Principal Component Analysis (KPCA) (Xu *et al.*, 2011), Kernel Fisher Discriminate Analysis (KFDA) (Mika *et al.*, 1999) and so on. It has been proved that KPCA and KFDA can perform well in extracting features compared with PCA.

More recently, the sparse representation technique has been widely used for face recognition. Sparse representation based classification (SRC) (Wright *et al.*, 2009) first represents a testing sample as a sparse linear combination of all the training samples via l_1 -norm minimization and then SRC classifies the testing sample by calculating which class leads to the minimum representation error. Sparse Preserving Projections (SPP) (Qiao *et al.*, 2010) is an unsupervised feature extraction based on sparse presentation, which aims at preserving the sparse reconstructive relations among samples in a low-dimensional space. It has been shown that SPP achieves better recognition results than other dimensionality reduction methods. However, using l_1 -norm (Donoho, 2006) to regularize sparse reconstructive coefficients is time-consuming. Local Sparse Representation Based Classification (LSRC) (Li *et al.*, 2010) is developed to free the computational burden of SRC by sparsely representing a testing sample from merely its neighbors. Zhang *et al.* (2011) proposes that it is the collaborative representation but not

the l_1 -norm sparse representation that makes SRC effective for pattern classification. It has been shown that using the non-sparse l_2 -norm to regularize the representation coefficients could lead to similar FR results to l_1 -norm regularization, but this can effectively reduce the computational complexity.

Motivated by the above analysis, locality and collaboration are combined in a more effective way in this study. Consequently, a novel feature extraction method named Kernel-based Collaborative Preserving Projection (KBCPP) is proposed for reducing dimension. As is well-known, most of previous face recognition methods usually exploit only the Euclidean distance to measure similarity between samples. However, the Euclidean distance couldn't reveal the internal distribution of some datasets. This indeed makes data inseparable. The kernel-induced distance measure (Chen and Zhang, 2004) is proposed for data separability. Thus, the proposed method KBCPP uses a kernel-induced distance measure to select K nearest neighbors of the target training sample from the remaining training samples. In KBCPP, a target training sample is collaboratively reconstructed from only its K nearest neighbors and a linear embedded space is calculated, where the local collaborative reconstructive relations among all the training samples are preserved. Experimental results on three benchmark databases show that the proposed method KBCPP has competitive classification results, but it has less complexity than SPP.

SPARSE REPRESENTATION

Let $X = [x_1, x_2, \dots, x_N]$ be the sample set of size N, x_i be the i th sample and x be the target sample. The sparse representation searches for a sparse coefficient vector s by:

$$\min \|s\|_0 \text{ s.t. } x = Xs \tag{1}$$

Equation 1 indicates that s is the coefficient vector and $\min \|s\|_0$ is the pseudo- l_0 norm which counts the number of non-zeros in s . Unfortunately, l_0 -minimization is NP-hard. It can be approximated by greedy algorithms (Mallat and Zhang, 1993) or corresponding l_1 -norm optimization problems.

SPARSE PRESERVING PROJECTION

Let $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{M \times N}$ be the training data matrix, x_i be the i th sample. SPP uses sparse representation to build graph G where $G = \{X, S\}$, $S = [s_1, s_2, \dots, s_N]$ is the edge weight matrix of the graph and is calculated by the l_1 -norm optimization problem:

$$\begin{aligned} &\min_s \|s_i\|_1 \\ &\text{s.t. } x_i = Xs_i \\ &\quad 1 = 1^T s_i \end{aligned} \tag{2}$$

where, s_i is the sparse coefficient vector of x_i , $1 \in \mathbb{R}^N$ is a vector of all ones. Then for best preserving the optimal weight, the objective function is obtained as follows:

$$\min_w \sum_{i=1}^n \|W^T x_i - W^T X s_i\|^2 \text{ s.t. } w^T X X^T w = 1 \tag{3}$$

Minimize Eq. 3 and calculate the projection matrix $W \in \mathbb{R}^{M \times M}$, where $M' < N$, keeping the optimal weight vector s_i at the same time. To avoid degenerate solutions, $w^T X X^T w = 1$ is commanded. Equation 3 can be written in a matrix form:

$$\max \frac{w^T X S_p X^T w}{w^T X X^T w} \tag{4}$$

Where:

$$S_p = S + S^T - S^T S$$

Therefore, the optimal W is the eigenvectors corresponding to the largest M' eigenvalues of the following generalized eigenvalue problem:

$$X S_p X^T w = \lambda X X^T w \tag{5}$$

KERNEL-BASED COLLABORATIVE PRESERVING PROJECTION

In sparse representation, the majority of the non-zero coefficients derive from neighbors of a target sample, namely a target sample can be expressed by its neighbors. The sparse coefficients obtained from l_1 -norm problem are time-consuming. Zhang *et al.* (2011) proposes that using the non-sparse l_2 -norm to regularize the representation coefficients could lead to similar FR results but can reduce computation complexity. Therefore, in order to improve recognition rate and reduce computation complexity of algorithm, KBCPP is proposed for feature extraction. KBCPP uses collaborative representation to reconstruct a target training sample from merely its neighbors. The KBCPP is described in detail as follows:

First, for each target training sample, its K nearest neighbors are determined from the rest of training samples by kernel-induced distance measure. Let $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{M \times N}$ be the training data matrix (N is the number of the training samples), x_i and $x_1, x_2, \dots, x_{i+1}, \dots, x_N$, respectively denote the target training sample and the rest

of training sample in the original space. Supposing there is a nonlinear function ϕ which can map an input space into a high dimensional feature space. Therefore, after mapping into the high dimensional feature space, the linearly inseparable problem changes into separable one. For each target training sample x_i , the K nearest neighbors are determined by the kernel-induced distance measure in the high dimensional feature space, the formula is shown as follows:

$$d(x_i, x_j) = \|\phi(x_i) - \phi(x_j)\|^2 \quad (i = 1, 2, \dots, N \quad j = 1, 2, \dots, i-1, i+1, \dots, N-1) \quad (6)$$

All the distances between the target training sample and the rest in the feature space can be calculated with Eq. 6, denoting them as d_1, d_2, \dots, d_{N-1} . Then the K nearest neighbors $x_{i_1}, x_{i_2}, \dots, x_{i_K}$ can be selected according to the distances. Here all the nearest neighbors of the target training sample can be defined as a matrix $X_{i\text{local}} = [x_{i_1}, x_{i_2}, \dots, x_{i_K}]$.

Then, the target training sample is reconstructed as a linear combination of its neighbors using collaborative representation. It assumes that the following equation is approximately satisfied:

$$x_i = a_{i_1}x_{i_1} + a_{i_2}x_{i_2} + \dots + a_{i_j}x_{i_j} + \dots + a_{i_K}x_{i_K} \quad (7)$$

where, x_{i_j} ($j = 1, 2, \dots, K$) are the K nearest neighbors of the i th target training sample and a_{i_j} ($j = 1, 2, \dots, K$) are the coefficients. The Eq. 7 can be rewritten into as follows:

$$x_i = X_{i\text{local}}\alpha_{i\text{local}} \quad (8)$$

Where:

$$X_{i\text{local}} = [x_{i_1}, x_{i_2}, \dots, x_{i_K}], \alpha_{i\text{local}} = [a_{i_1}, a_{i_2}, \dots, a_{i_K}]$$

In order to collaboratively represent the target training sample x_i using $X_{i\text{local}}$ with low computational burden, the regularized least square method is used. There is:

$$a_{i\text{local}} = \arg\min \{ \|x_i - X_{i\text{local}} \cdot a_{i\text{local}}\|_2^2 + \lambda \|a_{i\text{local}}\|_2^2 \} \quad (9)$$

where, λ is the regularization parameter. The regularization not only makes the least square solution stable, but also introduces a set number of ‘‘sparse’’ to the solution $\alpha_{i\text{local}}$, yet this ‘‘sparse’’ is slightly weaker than that by l_1 -norm. The solution $\alpha_{i\text{local}}$ of collaborative representation with regularized least square in Eq. 9 can be easily and analytically derived as:

$$a_{i\text{local}} = (X_{i\text{local}}^T X_{i\text{local}} + \lambda \cdot I)^{-1} X_{i\text{local}}^T x_i \quad (10)$$

Let $A = (X_{i\text{local}}^T X_{i\text{local}} + \lambda \cdot I)^{-1} X_{i\text{local}}^T$. Obviously, A is independent of x_i . Once a target training sample x_i comes, x_i can be projected onto A via Ax_i . This makes collaborative representation very fast and effective.

Without loss of generality, let $a_i = [a_{i_1}, a_{i_2}, \dots, a_{i_{N_i}}]^T \in \mathbb{R}^N$ be the collaborative coefficients of the reconstructed x_i and:

$$a_{i_j} = \begin{cases} a_{k\text{local}}, & x_k \text{ in } X_{i\text{local}} \text{ is } x_j \text{ in } X \\ 0, & \text{others} \end{cases} \quad (11)$$

The local collaborative representation of x_i is obtained, i.e. Xa_i .

Finally, like SPP, KBCPP tries to find the optimal linear projective vector v by:

$$\min_v \sum_{i=1}^N \|v^T x_i - v^T X a_i\|^2 \quad \text{s.t.} \quad v^T X X^T v = 1 \quad (12)$$

where, the total reconstructive errors are minimized. Equation 12 can be transformed into an equivalent maximization problem as follows:

$$\max \frac{v^T X (S^T + S - S S^T) X^T v}{v^T X X^T v} \quad (13)$$

where, $S = [a_1, a_2, \dots, a_{N_i}]$ by employing the Lagrange multiplier method, the solution v appears to be the eigenvectors of $(X X^T)^{-1} [X (S^T + S - S S^T) X^T]$ associated with largest eigenvalues.

ANALYSIS OF THE KERNEL-BASED COLLABORATIVE PRESERVING PROJECTION

In this section, the proposed method Kernel-based Collaborative Preserving Projection (KBCPP) is analyzed and compared with Sparse Preserving Projection (SPP). In SPP, to some extent, using l_1 -norm to regularize sparse reconstructive coefficients to represent the target training sample as a linear combination of the entire training sample can achieve good performance. However, the computational complexity is proportional to the scale of the problem, making the algorithm time consuming. To solve such drawback, a novel representation-based feature extraction method for face recognition method is proposed in this study. This algorithm can be seen as a sparse method. It uses l_1 -norm to regularize coefficients, intuitively reduces time complexity, consequently it can be much more efficient.

The KBCPP inherits the advantage of collaborative representation. What is improved is that selecting the nearest neighbors of the target training sample from all the training samples to represent the target training sample instead of all training samples. As a result, it can make full use of the information of neighbor samples, which leads to more precise classification.

In order to get better classification accuracy, the Radial Basis Function (RBF) is used when calculating the distance of the target training sample and the rest of the training samples in the feature space. As mentioned above, supposing that the nonlinear mapping φ is known, for each target training sample x_i in the original space, a new sample $\varphi(x_i)$ can be obtained in the high dimensional feature space. Hence, Eq. 6 can be extended as follows:

$$\begin{aligned}
 d(x_i, x_j) &= \|\varphi(x_i) - \varphi(x_j)\|^2 \quad (i=1, 2, \dots, N \quad j=1, 2, \dots, i-1, i+1, \dots, N) \\
 &= (\varphi(x_i) - \varphi(x_j))^T (\varphi(x_i) - \varphi(x_j)) \\
 &= \varphi(x_i)^T \varphi(x_i) - 2\varphi(x_i)^T \varphi(x_j) + \varphi(x_j)^T \varphi(x_j) \\
 &= K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j)
 \end{aligned}
 \tag{14}$$

Among the various kernel functions, Gaussian kernel function is used because of its better precise classification accuracy over the linear kernels on many samples. Therefore, in this study, the kernel function $K(x_i, x_j)$ is defined as follows:

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \sigma) \tag{15}$$

Clearly, for Gaussian kernel, $K(x, x) = 1$, hence, Eq. 14 can be written as follows:

$$\begin{aligned}
 d(x_i, x_j) &= \|\varphi(x_i) - \varphi(x_j)\|^2 \\
 &= 2(1 - K(x_i, x_j)) \\
 &= 2(1 - \exp(-\|x_i - x_j\|^2 / \sigma))
 \end{aligned}
 \tag{16}$$

As is well-known, in the original space, the Euclidean distance couldn't reveal the internal distribution of some datasets and the training samples become linearly inseparable. Therefore, the Euclidean distance can't deal with the distance between the target training sample and the rest of the training samples in complex situation, which can't select the really nearest neighbors. However, in the feature space, the training samples become linearly separable because of the use of nonlinear mapping. The kernel-induced distance measure can deal with the distance well even in complex situation. As a result, the K nearest neighbors determined by kernel-induced distance measure are close to the target training

sample. Moreover, the selected training samples can provide more information in reconstructing the target training sample, which leads to a better classification result.

EXPERIMENTAL RESULTS

In this section, a number of experiments were conducted to show the performance of Kernel-based Collaborative Preserving Projection (KBCPP), compared with Sparse Preserving Projection (SPP) (Qiao *et al.*, 2010). There are three face databases used in the experiment, namely ORL, Yale, AR face databases. The ORL face database consists images from 40 subjects; each subject provides 10 images with various facial expressions, varying lighting and facial details. The image size is cropped and resized to 92×112 . Yale database contains 15 individuals with 11 images for each person. Each face image is scaled to 80×80 with 256 gray levels. All the images are taken under varying facial expressions and lighting conditions. For the AR database, there are over 4000 images, which come from 126 subjects with each subject providing 26 images. In the experiment, a subset of the AR face database is used as samples which contains 100 subjects with the first 13 images selected per subject, and the image size is cropped to 80×100 . Before implementing all the methods, each sample vector is converted into a unit vector with the length of 1 in advance.

In the experiment, all the samples in each face database should be firstly divided into training and testing samples. For each of the ORL, Yale and AR databases, if m samples of all the n samples per class are selected for training samples and the remaining samples for testing samples, there are:

$$C_n^m = \frac{n(n-1) \cdots (n-m+1)}{m!}$$

possible combinations, as a result, there are C_n^m training sets and C_n^m testing sets for conducting experiments. For the ORL database, six images of each subject are used as the training samples and the remaining as the testing sample. As a result, different methods are tested by using 210 training and testing sample sets from the ORL database. For the Yale database, nine images of each subject are used as the training samples and there are 55 training and testing sample sets. As the AR database contained too many samples, the first seven samples from each class are selected as the training samples and the others as the test samples.

Sparse Preserving Projection (SPP) and Kernel-based Collaborative Preserving Projection (KBCPP) involve a PCA phrase to avoid the problem of singularity. In KBCPP, the distance between the target training sample and the rest of training samples in the feature space is calculated by Eq. 16, where σ is a positive constant and is set to 0.1 in the experiment. The collaborative coefficient a_{local} is solved by Eq. 10, where λ is set to 0.01. The nearest neighbor classifier is used to perform classification in all of the SPP and KBCPP experiments.

Figure 1-3 respectively show the recognition rates of the proposed method KBCPP versus the neighbors K on the ORL, Yale, AR databases. It can be seen that the recognition rates first increase then decrease with

variation of neighbor samples K. When the number of neighbor samples of the target training sample is small, the related class which the target training sample belongs to may not include its neighbors, leading to low recognition rate. With the increase of the number of neighbors, the recognition rates increase gradually. However, when the number of neighbors exceeds a threshold, there are a lot of disturbed samples in neighbors leading to the decrease of the recognition rates. From Fig.1, it can be seen that the proposed KBCPP can achieve the highest recognition rate when neighbor K is set to 130. For Fig. 2, the KBCPP method can obtain the highest recognition rate when neighbor K is set to 60. Figure 3 shows the KBCPP can achieve the best recognition rate when neighbor K is set to 150.

Figure 4 shows the experiment results of KBCPP and SPP on the ORL database. In the experiment, the value of K is set to 130, the feature dimension is set to 100. for each subject in the ORL database, 6 images are selected as the training samples. As mentioned above, there are 210 different combinations. Fig 4 shows the mean recognition rates of the different combinations. For each combination, the recognition rates obtained by the KBCPP are higher more or less than the algorithm SPP. Figure 5 shows the results of KBCPP and SPP on the Yale database when choosing 9 training samples and setting K = 60 and the feature dimension is set to 55. Similar conclusion to that in the previous experiments can be obtained from Fig. 5.

Figure 6-8 respectively demonstrates the curves of recognition rates of KBCPP and SPP versus the dimension on ORL, Yale, AR databases. For Fig.6, on the ORL database it can be seen that the recognition rates of KBCPP and SPP first increase then decrease with variation of dimension. When the dimension of feature is small, it

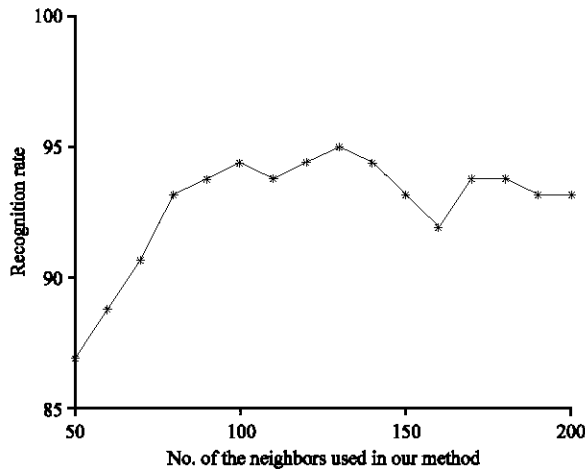


Fig. 1: Recognition rate of KBCPP versus neighbors K on ORL database

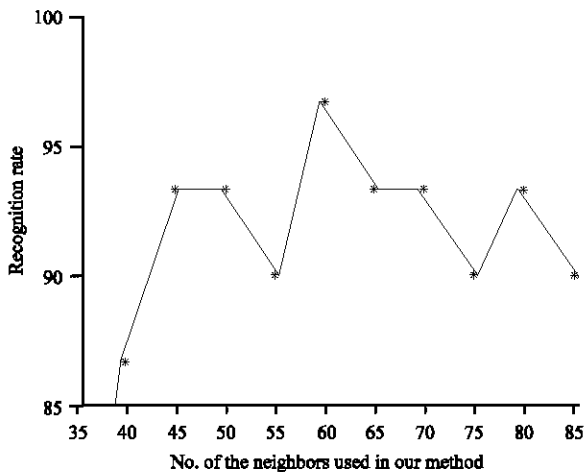


Fig. 2: Recognition rate of KBCPP versus neighbors K on Yale database

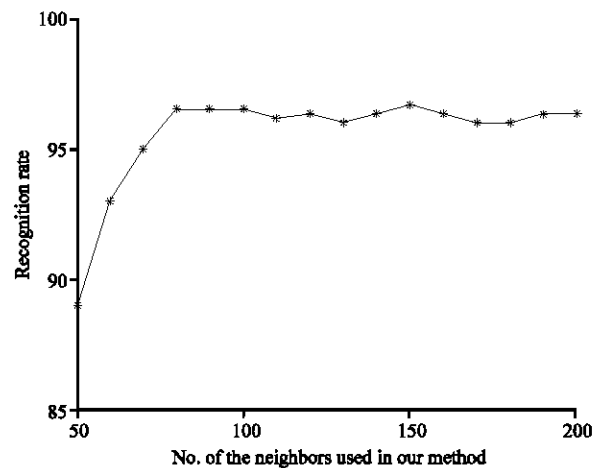


Fig. 3: Recognition rate of KBCPP versus neighbors K on AR database

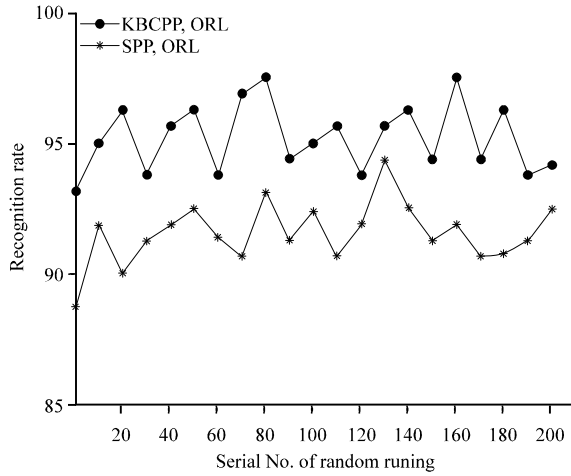


Fig. 4: Recognition rates of different combination by different methods on ORL database

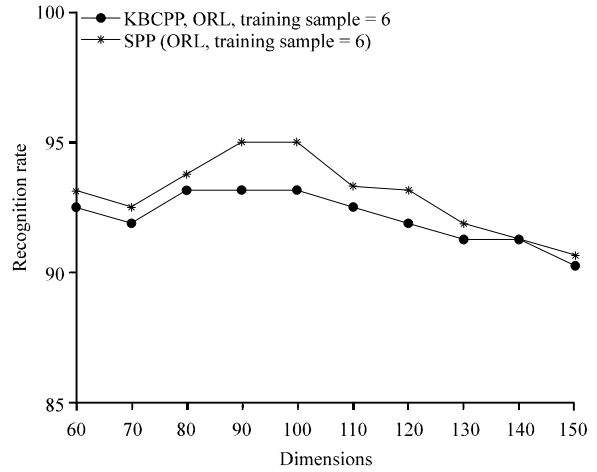


Fig. 6: Recognition rates by different methods versus feature dimension on ORL database

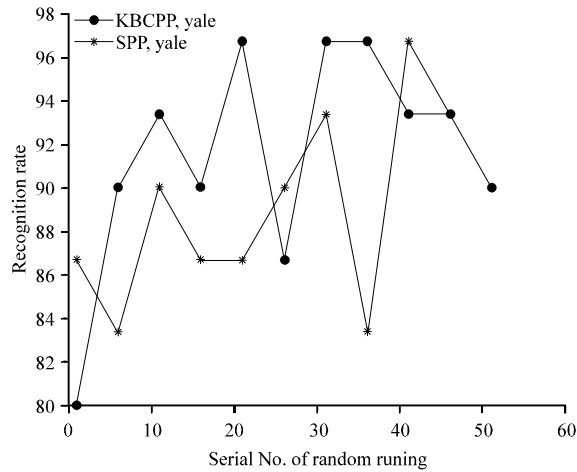


Fig. 5: Recognition rates of different combination by different methods on Yale database

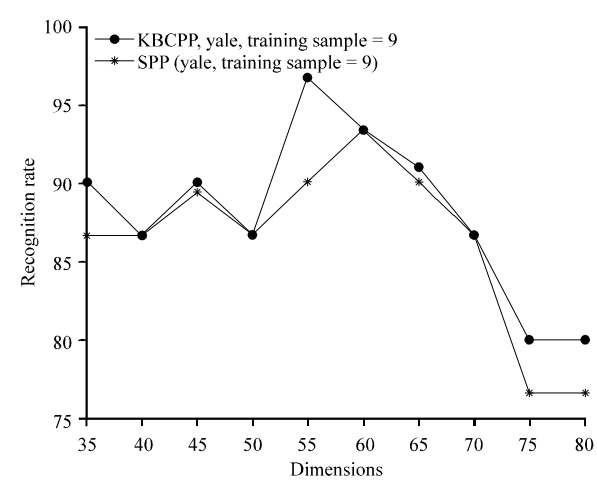


Fig. 7: Recognition rates by different methods versus feature dimension on Yale database

Table 1: The highest recognition rates of different methods on the ORL, Yale, AR database and the associated dimension of features

Face database						
	ORL		Yale		AR	
Feature	KBCPP	SPP	KBCPP	SPP	KBCPP	SPP
Rate (%)	95	93.125	96.667	93.333	97	96.333
	(k = 130)		(k = 60)		(k = 150)	
Dimension	100	100	55	60	80	120

leads to low recognition rate. This is because these features carry little information and can't effectively represent face image. With the increase of dimension, the recognition rates increase gradually. However, when the dimension of feature is too large, the recognition rates of KBCPP and SPP go into decline. This is because these features introduce interference. On the Yale database,

similar conclusion to that in the previous experiments can be obtained from Fig. 7. In Fig. 8, due to the over-fitting of features, the recognition rates of KBCPP and SPP change small when the dimension of feature is large. Generally speaking, from Fig. 6-8, it can be seen that the recognition rates of the proposed KBCPP method are higher than that of SPP in all dimensions. The maximal recognition rates of each method on ORL, Yale, AR databases with associated feature dimension are listed in Table 1, it shows that the proposed KBCPP method can reach the highest rate (95%) with least dimension (100), 96.667% with least dimension (55) and 97% with least dimension (80) on ORL, Yale, AR databases, respectively. However, the algorithm SPP separately reaches the

Table 2: Comparison of recognition rate and runtime

Feature	Recognition rate (%)			Runtime (sec)		
	ORL	Yale	AR	ORL	Yale	AR
KBCPP	93.125 (d = 100, k = 130)	93.333 (d = 55, k)	94.833 (d = 80, k = 150)	12.422	1.438	41.969
SPP	91.235 (d = 100)	90 (d = 55)	91.167 (d = 80)	21.657	6.953	247.234

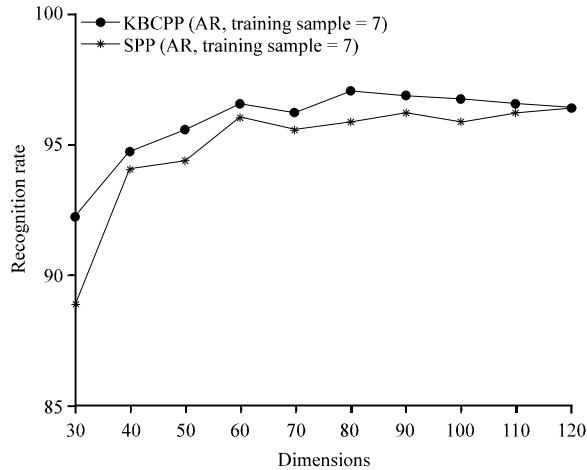


Fig. 8: Recognition rates by different methods versus feature dimension on AR database

highest rate (93.125%) with least dimension (100), 93.333% with least dimension (60) and 96.333% with least dimension (120) on ORL, Yale, AR databases.

Table 2 shows the mean recognition rate and runtime of all compared methods on the ORL, Yale, AR database. In Table 2 d and k respectively represents the appropriate dimension and number of neighbors when achieved best recognition rate. Compared with SPP, the proposed KBCPP separately improves the average recognition rates at least by 1.89% (=93.125-91.235%), 3.333% (=93.333-90%) and 3.666% (=94.833-91.167%) on ORL, Yale, AR database and the runtime of KBCPP is less than that of SPP, because KBCPP uses neighbors of target training sample to collaboratively represent target training sample, while SPP needs to calculate the global l_1 -sparse solutions using all samples.

CONCLUSION

In this study, a novel Kernel-based collaborative preserving projection is proposed. The method first selects k nearest neighbors for the target training sample from the rest of training samples using kernel-induced distance function and then collaboratively reconstructs the target training sample from its neighbors. Finally, the

method seeks a low-dimensional subspace where such local collaborative reconstructive relations among all training samples are preserved. Experimental results on ORL, Yale, AR face databases demonstrate that KBCPP has very competitive FR accuracy but with significantly lower complexity, compared with SPP.

REFERENCES

Belhumeur, P.N., J.P. Hespanha and D.J. Kriegman, 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intel.*, 19: 711-720.

Chen, S. and D. Zhang, 2004. Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. *IEEE Trans. Syst. Man Cybernetics*, 34: 1907-1916.

Donoho, D.L., 2006. For most large under determined systems of linear equations the minimal₁-norm solution is also the sparsest solution. *Comm. Pure Applied Math.*, 59: 797-829.

He, X., S. Yan, Y. Hu, P. Niyogi and H.J. Zhang, 2005. Face recognition using Laplacian faces. *IEEE. Trans. Pattern Anal. Machine Intell.*, 27: 328-340.

Jing, X.Y., D. Zhang and Y.Y. Tang, 2004. An improved LDA approach. *IEEE Trans. Syst. Man. Cybern. Part B*, 34: 1942-1951.

Li, C.G., J. Guo and H.G. Zhang, 2010. Local sparse representation based classification. *Proceedings of the 20th International Conference on Pattern Recognition*, August 23-26, 2010, Istanbul, pp: 649-652.

Mallat, S.G. and Z. Zhang, 1993. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41: 3397-3415.

Mika, S., G. Ratsch, J. Weston, B. Scholkopf and K.R. Mullers, 1999. Fisher discriminant analysis with kernels. *Proceedings of the IEEE Neural Networks for Signal Processing Workshop*, August 1999, Madison, WI., pp: 41-48.

Qiao, L., S. Chen and X. Tan, 2010. Sparsity preserving projections with applications to face recognition. *Pattern Recogn.*, 43: 331-341.

- Roweis, S.T. and L.K. Saul, 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290: 2323-2326.
- Wright, J., A.Y. Yang, A. Ganesh, S.S. Sastry and Y. Ma, 2009. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31: 210-227.
- Xu, Y., D. Zhang, J. Yang, Z. Jin and J.Y. Yang, 2011. Evaluate dissimilarity of samples in feature space for improving KPCA. *Int. J. Inform. Technol. Decision Making*, 10: 479-495.
- Zhang, L., M. Yang and X. Feng, 2011. Sparse representation or collaborative representation: Which helps face recognition? *Proceedings of the IEEE International Conference on Computer Vision, IEEE*, November 6-13, 2011, Barcelona, Spain, pp: 471-478.