Copyright ⓒ2016 KSII

An improved kernel principal component analysis based on sparse representation for face recognition

Wei Huang, Xiaohui Wang, Yinghui Zhu, Gengzhong Zheng
School of Computer and Information Engineering, Hanshan Normal University,
Qiaodong, Chaozhou, China
[e-mail: hw@hstc.edu.cn]
*Corresponding author: Wei Huang

Received November 10, 2015; revised February 19, 2016; accepted March 26, 2016; published June 30, 2016

Abstract

Representation based classification, kernel method and sparse representation have received much attention in the field of face recognition. In this paper, we proposed an improved kernel principal component analysis method based on sparse representation to improve the accuracy and robustness for face recognition. First, the distances between the test sample and all training samples in kernel space are estimated based on collaborative representation. Second, S training samples with the smallest distances are selected, and Kernel Principal Component Analysis (KPCA) is used to extract the features that are exploited for classification. The proposed method implements the sparse representation under ℓ_2 regularization and performs feature extraction twice to improve the robustness. Also, we investigate the relationship between the accuracy and the sparseness coefficient, the relationship between the accuracy and the dimensionality respectively. The comparative experiments are conducted on the ORL, the GT and the UMIST face database. The experimental results show that the proposed method is more effective and robust than several state-of-the-art methods including Sparse Representation based Classification (SRC), Collaborative Representation based Classification (CRC), KCRC and Two Phase Test samples Sparse Representation (TPTSR).

Keywords: Kernel PCA,Sparse Representation, Collaborative Representation, Small Sample Size Problem, Face Recognition

http://dx.doi.org/10.3837/tiis.2016.06.014 **ISSN : 1976-7277**

1. Introduction

Face recognition has received considerable attention in pattern recognition and machine learning because of its wide applications [1, 2]. In the past several decades, a large number of outstanding face recognition methods have been proposed and face recognition has made much progress. However, face recognition is still a great challenge owing to the following two main reasons. First, the face image matrices are always transformed into high-dimensional vectors in the face recognition problem. The image matrix-to-vector transformation will cause the rapid incensement of the dimensionality. This is so called "curse of dimensionality" [3]. Compared with the dimensionality of high-dimensional data, the number of the training samples is much less. The high-dimensional Small Size Sample (SSS) problem will lead to the over fitting or unstable problems in face recognition [4]. Second, the face images of one individual may contain variations of illuminations, postures and expressions. This means that the data uncertainty in face recognition may severely affect the performance of face recognition [5]. In general, the training sample set would not include all the varieties. Consequently, insufficient training samples will lead to low accuracies [6-8]. Furthermore, if the difference between different individuals is much smaller than it in the same individual , which is caused by the variations, the performance will be dropped sharply. To add virtual face images to the set of training samples is partially able to remedy the difficulty [9].

In order to overcome the "curse of dimensionality", feature extraction technology is applied in face recognition. Principal Component Analysis (PCA) [7] and Linear Discriminant Analysis (LDA) [8] are two of the greatest well-known dimension reduction techniques. PCA constructs a serial of orthogonal vectors to pursuit a best representation while LDA is able to separate different classes in the low-dimensional subspace. Up to now, the researchers continue to improve the performances of PCA and LDA since they were proposed [9-12, 51]. Two-dimensionality extension and kernel extension are two important aspects. Two-dimensionality PCA (2DPCA) [13] and Two-Dimensionality LDA (2DLDA) [14] not only preserve structure information embedded in face images, but also avoid the SSS problem and reduce the computational time. As the improvement of LDA, Uncorrelated Linear Discriminant Analysis (ULDA) [15] attracts much attention. Although these methods that are based on PCA and LDA have been applied successfully, they are linear transformation and sensitive to noise. As for high-dimensionality data such as the face images data, they are not always linear distribution in the space. In order to extract the non-linear characteristics, kernel technology is applied in PCA [16, 17] and LDA [18]. Furthermore, the researchers do many works to improve the performance of kernel PCA in face recognition in recent years. For example, Fang et al proposed kernel representation-based nearest neighbor classifier [19]. Lu utilized the symmetry of face image and proposed symmetrical principal component analysis (SPCA) to improve the accuracy [20], Heo et al use fuzzy memberships to extend the kernel PCA [21]. Reference [22] used a few partial data points to determine which data points can be used to reduce time consuming and computation memory in kernel PCA. Besides face recognition, kernel PCA is used for decentralized fault detection [23, 24], denoising [25], identification of severe weather outbreaks [26], similarity invariants shape recognition [27], microarray gene data analysis [28].

In recent years, the representation based classifications, such as Sparse representation based classification (SRC) [29] and collaborative representation based classification (CRC) [30] received much attention in face recogntion field. SRC searches a "sparse" linear combination of training samples to represent the test sample. In other words, SRC chooses a subset from

training set to represent test samples through ℓ_1 regularization. The experiment results show that SRC exceeds many other algorithms and is robust to occlusion, illumination and noise. However, ℓ_1 regularization causes the high computational cost because it is not a simple closed-form solution. After that, SRC with $\ell_{2,1}$ [31] regularization and ℓ_2 regularization [32, 33] were proposes to reduce the computational cost. Compared with conventional SRC, these proposed methods achieve satisfactory and robust face recognition results. Though conventional SRC usually leads to sparser, SRC with $\ell_{2,1}$ regularization can receive group sparsity [34, 35]. Moreover, SRC with $\ell_{2,1}$ is implemented using ℓ_1 and ℓ_2 regularizations simultaneously [35-38], which is different from conventional SRC. A lot of research works have been done to improve SRC. Xu et al proposed a supervised sparse representation method with a heuristic strategy [39]. Jiang et al proposed a semi-supervised discriminant analysis based sparse method for face recognition [40]. Reference [41] divided all the training samples into several blocks and then determined whether the block is in occluded using the linear regression technique. A two-phase test samples sparse representation method was proposed to reduce the high computational cost and improve the performance of the face recognition [42]. Lia et al proposed a new decision rule for sparse representation. Wang et al extent SRC to kernel space and proposed multi-kernel learning for sparse representation [43]. Reference [6] induced a kernel distance to determine N nearest neighbors of the test samples from training set to realize the "sparseness". Why can sparse representation receive the high accuracy and robustness for face recognition? Zhang et al in reference [30] attribute it to that sparse representation chooses from all the training samples to represent the test sample. In other words, SRC utilizes the similarity of face images to reduce the unreasonable representation residual. This strategy induces collaborative representation based classification (CRC). CRC can play a similar role to the sparse ℓ_0 regularization in enhancing the discrimination of representation.

However, if the test sample that has the same differential vector with two types of training samples, SRC cannot distinguish their categories. In order to prevent such problem, some researchers proposed a method which combine sparse and kernel technology to perform classification. In this paper, we propose an improvement approach for face recognition, which is based on sparse representation and PCA in kernel space. The proposed method estimates the distances between test sample and the all the training samples based on collaborative representation idea in kernel space. We select S training samples which get the S smallest distances. After that, KPCA is used to reduce the dimensionality and extract the most relevant information for classification. The proposed method implements the sparse representation under ℓ_2 regularization and extracts the most relevant information twice to improve the robustness. In order to test the performance of the proposed method, we compare it with several state-of-the-art methods including SRC, CRC, KCRC and TPTSR in the ORL, the GT and the UMIST face databases. The experiment results show that our method is more effective and robust.

The rest of this paper is organized as follows: Section 2 demonstrates the related works; Section 3 discusses our proposed method in detail; Section 4 conducts extensive experiments to demonstrate the performance of our works; and Section 5 concludes the paper.

2. Related works

In this section, some important works including SRC, TPTSR, CRC and Kernel CRC, which related to our works are reviewed. We suppose that we have c individuals with n_i training samples from the i^{th} individual, $i = 1, 2, \cdots, c$ and $n = \sum_{i=1}^{c} n_i$. The training set is $X =$

$[x_1, x_2, \cdots, x_n]$, X_i is the sub training set that contains the training samples from the i^{th} individual, and y is the test sample.

2.1 Sparse representation

The main assumption of sparse representation is that the training samples from one individual are lie on a subspace. So the test samples from the c^{th} class can be represented approximately by the linear span of X_c:

$$y = X_c \alpha_c \tag{1}$$

where α_c is the reconstructive coefficients. Eq. (1) can be rewritten in term of the training samples from all individuals as

$$y = X\alpha_0 \tag{2}$$

where $\alpha_0 = [0^T, (\alpha_c)^T, 0^T]$. Since c is unknown, SRC aims to solve the following ℓ_0 minimization problem:

$$(\ell_0): \hat{\alpha}_0 = arg\, min\|\alpha\|_{0,} \ subject\ to\ y = X\alpha \tag{3}$$

where $\|\cdot\|_0$ denotes the ℓ_0 norm, which is the number of non-zero entries in the vector.

Unfortunately, it is a NP-hard problem to solve $\hat{\alpha}_0$ even to get the approximate solution. According to compressive sensing [44, 45], if the $\hat{\alpha}_0$ is sparse enough the solution of ℓ_0 minimization problem is approximately equal to ℓ_1 minimization problem:

$$(\ell_1):\ \hat{\alpha} = argmin\|\alpha\|_1, \ \ subject\ to\ y = X\alpha \tag{4}$$

Considered the occlusion, Eq. (4) can be expressed as:

$$(\ell_1^s): \hat{\alpha} = arg\, min\|\alpha\|_1, \ subject\ to\ \|y - X\alpha\| \le \varepsilon \tag{5}$$

where $\varepsilon > 0$ is a given tolerance.

After that, the representative residual can be gotten by:

$$e_i = \|y - X_i\hat{\alpha}_i\|_2^2, \ i = 1,2,\cdots,c \tag{6}$$

where $\hat{\alpha} = [\hat{\alpha}_1, \cdots \hat{\alpha}_i, \cdots, \hat{\alpha}_c]^T$, $\hat{\alpha}_i$ is the coefficient from X_i. Finally, the test sample will be identified by:

$$identity(y) = arg_i\, min\,\{e_i\} \tag{7}$$

2.2 Collaborative Representation

In order to reduce the unreasonable representation residual, all the training samples are used to represent the test sample via a linear combination. Compared with SRC, CRC receives not only high performance, but also low computational cost. CRC aims to solve the following ℓ_2 minimization problem:

$$\left(\hat{\xi}\right) = argmin_{\xi}\{\|y - \xi X\|_2^2 + \lambda\|\xi\|_2^2\} \tag{8}$$

where λ is the regularization parameter. λ makes the least square solution much more stable and a certain amount of "sparsity" to $\hat{\xi}$.

The least squares estimation is performed to estimate the coefficient:

$$\hat{\xi} = (X^T X + \lambda I)^{-1} X^T y \tag{9}$$

The representation residual e_i can be calculated as follows:

$$e_i = \left\|y - \sum_{p=1}^{n_i} \hat{\xi}^{p,i} X^{p,i}\right\|_2^2,\ i = 1,2,\cdots,c \tag{10}$$

where $\xi^{p,i}$ and $X^{p,i}$ represent the p^{th}coefficient and the p^{th} training sample from the i^{th} class respectively.

The rule in favor of the class with the minimum distance can be calculated by:

$$identity(y) = argmin_i\{e_i\} \tag{11}$$

2.3 Kernel Collaborative Representation

Kernel method is the effective technology to extract the nonlinear feature and has been applied in computer vision and pattern recognition in recent years [46, 47]. If we use a nonlinear mapping φ, the original data space $\mathcal{R}$ is mapped into a higher dimensional feature space $\mathcal{F}$

$$\varphi: \mathcal{R} \to \mathcal{F}$$
$$X \to \varphi(X)$$

Denote the mapped samples from the original feature space as $\phi = [\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_N)]$. The objective function of Kernel CRC can be written as:

$$\psi = argmin_{\psi}\{\|\varphi(y) - \phi\psi\|_2^2 + \lambda\|\psi\|_2^2\} \tag{12}$$

where λ is the regularized parameter. It is easy and analytical to solve $\hat{\psi}$ as

$$\hat{\psi} = (\phi^T\phi + \lambda I)^{-1}\phi^T\varphi(y) \tag{13}$$

Suppose that there is a kernel $k(\cdot,\cdot)$ induced by feature mapping function φ and $k(x_i, x_j) = \varphi^T(x_1)\varphi(x_2)$ represents a nonlinear similarity between two vectors x_i and x_j.

$$\phi^T\phi = [\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_N)]^T \cdot [\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_N)] \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \vdots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix}$$

(14)

$$\phi^T \cdot \varphi(y) = [\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_N)]^T \cdot \varphi(y) = \begin{pmatrix} k(x_1, y) \\ \vdots \\ k(x_n, y) \end{pmatrix} \tag{15}$$

If we denote $K = \begin{pmatrix} k(x_1, x_1) & \ldots & k(x_1, x_n) \\ \vdots & \vdots & \vdots \\ k(x_n, x_1) & \ldots & k(x_n, x_n) \end{pmatrix}$, $k_y = \begin{pmatrix} k(x_1, y) \\ \vdots \\ k(x_n, y) \end{pmatrix}$

We can use Eq.(16) to calculate the coefficients:

$$\hat{\psi} = (K + \lambda I)^{-1} k_y \tag{16}$$

Then we can calculate the representation residual for each class by

$$e_i = \left\| k_y - \sum_{p=1}^{n_i} \hat{\psi}^{p,i} \, \hat{k}^{p,i} \right\|_2^2, \; i = 1,2,\cdots,c \tag{17}$$

where n_i is the number of the training samples from the i^{th} class, $\hat{\psi}^{p,i}$ and $\hat{k}^{p,i}$ represent the p^{th} coefficients and the p^{th} training samples from the i^{th} class in kernel space respectively. Finally, we identify the class which makes the minimums representation residual.

$$identity(y) = arg\ min_i\{e_i\} \tag{18}$$

2.4 Two-phase test samples sparse representation

The main assumption of TPTSR is that the test sample and its *M* "nearest neighbors" are probably from the same class. First, TPTSR searches *M* "nearest neighbors" for test sample via CRC. Second, TPTSR represents the test sample via linear combination of its *M* "nearest neighbors". At last, the test sample is determined by the representation residual. Compared with CRC and SRC, TPTSR ignores the training samples that are far from the test sample.

SCR implements the sparse ℓ_0 regularization. While CRC enhances the discrimination of representation through ℓ_2 regularization. KCRC extents CRC to kernel space to improve the nonlinear classification ability. TPTSR searches *M* "nearest neighbors" of test sample for CRC. We combine the merits of TPTSR and KPCA to propose an improved kernel principal component analysis method based on sparse representation to improve the accuracy and robustness for face recognition.

3. The proposed method

3.1 Presentation of the proposed method

In this section, we present the improved kernel principal component analysis method for face recognition. We first search the *S* nearest training samples of the test sample in kernel space. And then the features are extracted by using KPCA. At last, these features are used to identify

the class label of the test sample. The specific steps are as follows:

Step 1 Search S nearest training samples in kernel space

We first assume that all the samples are mapped into a new space by using mapping φ. It is also supposed that in kernel space the test samples y can be approximately represented by the combination of all training samples:

$$\varphi(y) = \sum_{i=1}^{n} w_i \varphi(x_i) \tag{19}$$

where w_i is the representation coefficient of the i^{th} training sample. Using the least squares algorithm, we can obtain the following solution:

$$W = (K^T K + \gamma I)^{-1} K^T K_y \tag{20}$$

where

$W = [w_1, w_2, \cdots, w_n]^T$, $K = \begin{bmatrix} k_{1,1} & \cdots & k_{1,n} \\ \vdots & \cdots & \vdots \\ k_{n,1} & \cdots & k_{n,n} \end{bmatrix}$, $k_{i,j} = k(x_i, x_j) = \left(\varphi(x_i), \varphi(x_j)\right)$, $K_y = X^T \varphi(y) = \left(k(y, x_1), k(y, x_2), \cdots, k(y, x_n)\right)^T$, I is the identity matrix and γ is a small positive constant makes the least square solution much more stable. $\left(\varphi(x_i), \varphi(x_j)\right)$ stands for the inner product of $\varphi(x_i)$ and $\varphi(x_j)$. $X = [\varphi(x_1) \ldots \varphi(x_n)]$. In order to search S nearest training samples of the test sample, we give the distance defined as follows:

$$q_i = \|\varphi(y) - \omega_i \varphi(x_i)\|_2^2 = k(y, y) + (\omega_i)^2 k(x_i, x_i) - 2\omega_i k(y, x_i), \quad i = 1,2, \cdots, n \tag{21}$$

From Eq. (21) we know that the smaller q_i is, the higher relevance between the test sample and the training sample is. So we are able to select the S nearest training samples based on q_i. The samples which get the S smallest distances are selected and denoted as $X' = [x'_1, x'_2, \cdots, x'_S]$ and the number of the i^{th} individual samples is denote as $\hat{n}_i$ and $S = \sum_{i=1}^{c} \hat{n}_i$.

Step 2 Feature extraction using KPCA.

In this step, we want to extract the feature from X'. Similar with step 1, we have

$k'_{i,j} = k(x'_i, x'_j) = \left(\varphi(x'_i), \varphi(x'_j)\right)$, $K' = \begin{bmatrix} k'_{1,1} & \cdots & k'_{1,s} \\ \vdots & \cdots & \vdots \\ k'_{s,1} & \cdots & k'_{ss} \end{bmatrix}$,

$K'_y = \left(k(y, x'_1), k(y, x'_2), \cdots, k(y, x'_S)\right)^T$.

Then the eigenvalues λ and eigenvectors v of covariance matrix C are given by

$$K'v = \lambda v \tag{22}$$

We select the first d eigenvectors which correspond to the d largest eigenvalues. They are denoted as $\hat{A} = [az_1, az_2, \cdots, az_s]^T$ and $C = [\lambda_1, \lambda_2, \cdots, \lambda_d]$ respectively. az_i is the feature we extract from the training sample $\varphi(x'_i)$. This means $[az_1, az_2, \cdots, az_s]^T = K'C$. Similarly,

we get the feature of the test sample $\varphi(y)$, which is denoted as $b\varphi(y)$ by

$$[b\varphi(y)]^T = \left(K'_y\right)^T C \tag{23}$$

Step 3 Classification

In this step we will calculate the distance between $b\varphi(y)$ and az_i, where $i = 1,2,\cdots,c$. Then the test sample will be identified to the class which it get the smallest distance.

First, the S nearest training samples are used to represent sparsely the test samples. The represent coefficients are estimated by:

$$\hat{F} = \left(\hat{A}^T\hat{A} + \gamma I\right)^{-1}\hat{A}^T\hat{B} \tag{24}$$

where $\hat{B} = b\varphi(y)$, $\hat{F} = \left[\hat{f}_1^1, \hat{f}_1^2, \cdots, \hat{f}_i^{\hat{n}_i}, \cdots \hat{f}_c^{\hat{n}_c}\right]^T$, γ is the regularization parameter and makes the least square solution much more stable. After that, the residual can be calculated by

$$d_i = \left\|\hat{B} - \sum_{j=1}^{\hat{n}_i} \hat{f}_i^j\left(az_i^j\right)\right\|_2^2, i = 1,2,\cdots,c \tag{25}$$

where $\hat{f}_i^j$ is the i^{th} class the j^{th} sample coefficient from $\hat{F}$ and az_i^j is the i^{th} class the j^{th} feature from $\hat{A}$.

Finally, the classification rule in favor of the class with the minimum residual can be expressed as

$$identify(y) = argmin_i\{d_i\} \tag{26}$$

3.2 Analysis

There are several advantages in our method. First, like the conventional kernel method, the proposed method is able to increase the separability of the samples in kernel space. This means that the distances between different classes will be enlarged when the kernel trick is used. The theoretical and experiment results offered in [12] show that the samples that are mapped into the higher-dimensional space is more beneficial to correct classification. Second, the proposed method extracts the most relevant features for test samples twice: (1) we select the nearest training samples globally in kernel space. All the samples including the test sample are mapped into the kernel space. At the same time, the collaborative representation idea is used to select the most nearest samples in the new feature space. (2) KPCA is implemented to extract dominating features from the test sample and S nearest training samples. The sparse representation of the test sample is executed with ℓ_2 regularization. So the extracted features are more suitable for classification.

4. Experimental Classification Results and Analysis

In this section, we compare our method with several stat-of-the-art face recognition methods including SRC, CRC, KCRC and TPTSR on the ORL [48], the GT [49] and the MUIST [50] face databases. In our experiments, all the face images were normalized to 32×32 pixel. The first $t(\ t = 2, 3, 4, 5, \cdots)$ samples per individual are used for training and the remaining samples are used for test. In our method, KPCA is implemented to reduce the dimension and contain 95% energy.

4.1 Experiments on ORL face database

The ORL face database consists of a total of 400 images from 40 people. Each person has 10 images. For some individuals, the images are taken at different times. The varieties of this face database include the open or closed eyes, smiling or no smiling, wearing glasses or no glasses. Also the face images are taken with a tolerance for some titling and rotation of the face of up to 20^o. And there are some variations in the scale of up to about 10%. **Fig. 1** shows some samples of the ORL face database.

Fig. 1. All the images from one individual on ORL face database.

In this experiment, we choose the first t ($t = 2,3,4,5,6$) of each individual to form the training set and the remaining samples for test. The total number of training samples is different according to different t. So we set S to be an integer. Suppose s is the sparseness coefficient and S equal s multiply to the total number of training samples. In all experiments, s is set as 0.6. The comparative results are summarized in **Table 1**.

Table 1. Comparative result of different methods on ORL face database

Method	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
SRC	0.8250	0.8571	0.9083	0.8850	0.9187
CRC	0.7656	0.8571	0.9250	0.9400	0.9187
KCRC	0.8844	0.8929	0.9125	0.9250	0.9437
TPTSR	0.8344	0.8827	0.8958	0.8900	0.9375
the proposed method	0.8844	0.9000	0.9250	0.9200	0.9437

From **Table 1**, we can find that the two methods based on kernel, KCRC and the proposed method, get the highest accuracies. When $t = 2$ and $t = 6$ KCRC and our method have the same accuracies. When $t = 3$ and $t = 4$, our method is better than any other methods including KCRC. In general, the proposed method is efficient.

In order to investigate the relationship between accuracy and the coefficients, s and dimensionality, we estimate the accuracies by fixing s and dimensionality respectively in different number of training samples. **Fig. 2** shows the accuracy variation with different sparseness coefficient s (s is setting from 0.1 to 1.0 and the step is 0.1) when PCA contains 100% energy. **Fig. 3** shows the relationship between accuracy and dimensionality by fixing the sparseness coefficient ($s = 0.6$). **Table 2** shows the highest accuracy and the sparseness coefficient s in each number of training samples.

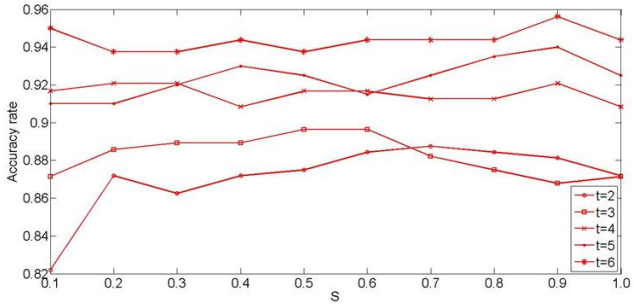


Fig. 2. The accuracies in different sparseness coefficient s (from 0.1 to 1.0) in the ORL face database.

From **Fig. 2** we can find that the accuracies seem to oscillate according in different sparseness coefficient s. The accuracy cannot reach the highest level when all the training samples are used ($s = 1$). Though face recognition is a Small Sample Size Problem, some training samples are not suitable for classification.

Table 2. The highest accuracy and its value of sparseness coefficient (s) in ORL face database

Item	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
Accuracy	0.8875	0.8964	0.9208	0.9400	0.9563
s	0.7	0.5	0.2	0.9	0.9

We can find from **Table 2** that the highest accuracy in different number of training samples need different sparseness coefficient. Also, the accuracies when $t = 2$ ($s = 0.7$) and $t = 6$ ($s = 0.9$) in **Table 2** are higher than those in **Table 1**. This means that the dimension and sparseness coefficient s are all important for the accuracy. In order to investigate the relationship between accuracy and dimension, we calculate the accuracies according to different dimension by fixing the sparseness coefficient ($s = 0.6$). **Fig. 3** and **Table 3** show the experiment results.

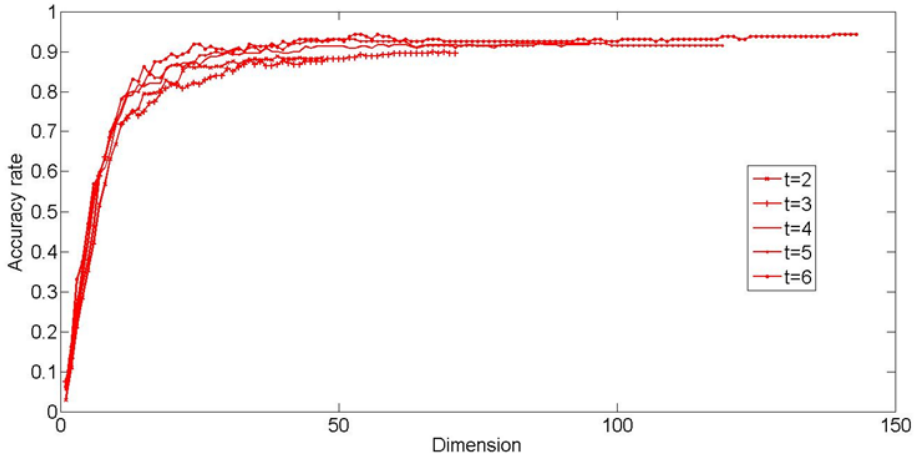


Fig. 3. The accuracies in different dimension in ORL face database.

From **Fig. 3** we can observe that the accuracies increase sharply in any cases when the dimensionality is small. When the dimensionality is large enough the accuracies oscillate according to the increment of dimensionality. **Table 3** The highest accuracy and the dimension in ORL face database.

Table 3. The highest accuracy and the dimension in ORL face databas

Item	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
Accuracy	0.8875	0.9000	0.9208	0.9300	0.9437
Dimensionality	39	67	82	49	53

Table 3 shows that the highest accuracy does not happen in the largest dimensionality. From the experiment results that shown in section 4.1, we can find that if the suitable sparseness coefficient and dimensionality are set, the higher accuracies can be gotten.

4.2 Experiments on GT face database

The Georgia Tech (GT) face database contains images of 50 people taken two or three sessions. The pictures show frontal and/or tilted faces with different facial expressions, lighting conditions and scale. Each image is manually labeled determine the position of the face in the image. **Fig. 4** shows all the images of one subject on GT face database.

Fig. 4. All the images of one individual on Georgia Tech face image dataset.

In this experiments, we choose the first t ($t = 2, 3, 4, \cdots, 14$) samples per individual to form the training set and the remaining samples are used for test. The contrastive methods include SRC, CRC, KCRC and TPTSR. The experiment results are shown in **Table 4**.

Table 4. Comparative result of different methods on GT face database

Method	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$	$t = 8$	$t = 9$	$t = 10$	$t = 11$	$t = 12$	$t = 13$	$t = 14$
SRC	0.4154	0.4517	0.4727	0.5100	0.5756	0.6175	0.6543	0.6533	0.6520	0.7200	0.7200	0.8100	0.8200
CRC	0.3892	0.4083	0.4055	0.4100	0.4467	0.4975	0.4886	0.5000	0.4760	0.4950	0.4600	0.4900	0.4800
KCRC	0.4554	0.4933	0.5309	0.5540	0.6333	0.685	0.7114	0.7367	0.7640	0.8000	0.8267	0.8600	0.8600
TPTSR	0.4677	0.5200	0.5673	0.5920	0.6489	0.6850	0.7286	0.7333	0.7600	0.7900	0.7933	0.8000	0.8200
the proposed method	**0.5000**	**0.5433**	**0.5909**	**0.6120**	**0.6844**	**0.7225**	**0.7314**	**0.7800**	**0.7880**	**0.8300**	**0.8533**	**0.8900**	**0.9000**

From **Table 4** we can observe that the methods based on kernel have a big improvement than the other methods. Though CRC get the lowest accuracy, KCRC is able to get higher accuracies than SRC. On the GT face database, TPTSR which is under ℓ_2 regularization is better than SRC, which is under ℓ_1 regularization. We owe the improvement to the tactics which selects M nearest neighbors for the test sample. The proposed method not only inherit this merit, but also extent it to the kernel space. From Fig. 4 we can observe that the face images are under big variations. These variations will bring the difficulty fo feature extraction and classification. So we can find some methods such as CRC and SRC can not get the high

accuracy in Table 4. The proposed method gets higher accuracies than any other methods in any causes. Compared with KCRC, our method has a big improvement. The average accuracy of KCRC and our method are 68.52% and 72.51% respectively. This means the proposed method has over 5.8% improvement. The experimental results show that the proposed method is effective and more robust than the others.

In order to investigate the relationship between accuracy and sparseness coefficient, the relationship between accuracy and dimensionality respectively, we calculate the accuracies with different sparseness coefficients by fixing dimensionality and calculate the accuracies with different dimensionalities by fixing sparseness coefficients. The experimental results are shown in **Fig. 5**, **Table 5** and **Fig. 6**, **Table 6** respectively.

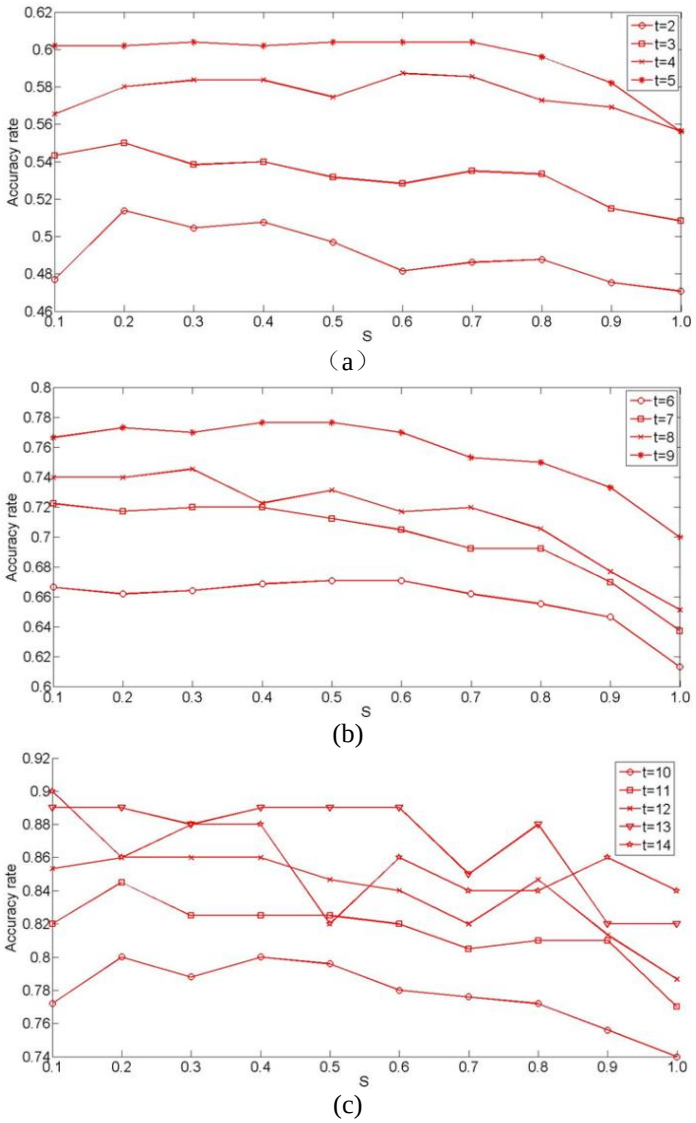


Fig. 5. The accuracies in different sparseness coefficient s (from 0.1 to 1.0) in GT face database. (a), (b) and (c) are the result with different number of training sample.

Table 5. The highest accuracy and its value of sparseness coefficient (s) in GT face database

Item	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$	$t = 8$	$t = 9$	$t = 10$	$t = 11$	$t = 12$	$t = 13$	$t = 14$
Accuracy	0.5138	0.5500	0.5873	0.6040	0.6711	0.7225	0.7457	0.7767	0.8000	0.8450	0.8600	0.8900	0.9000
s	0.2	0.2	0.6	0.3	0.5	0.1	0.3	0.4	0.2	0.2	0.2	0.1	0.1

From **Fig. 5** we can see that almost all the accuracies hit rock bottom when all the training samples are used for classification. We can observe from **Table 5** that at most situations the highest accuracies happen in high sparseness level in the GT face database.

Fig. 6 and **Table 6** show relationship between accuracy and dimensionality. We calculate the accuracies according to different dimension by fixing the sparseness coefficient ($s = 0.6$).

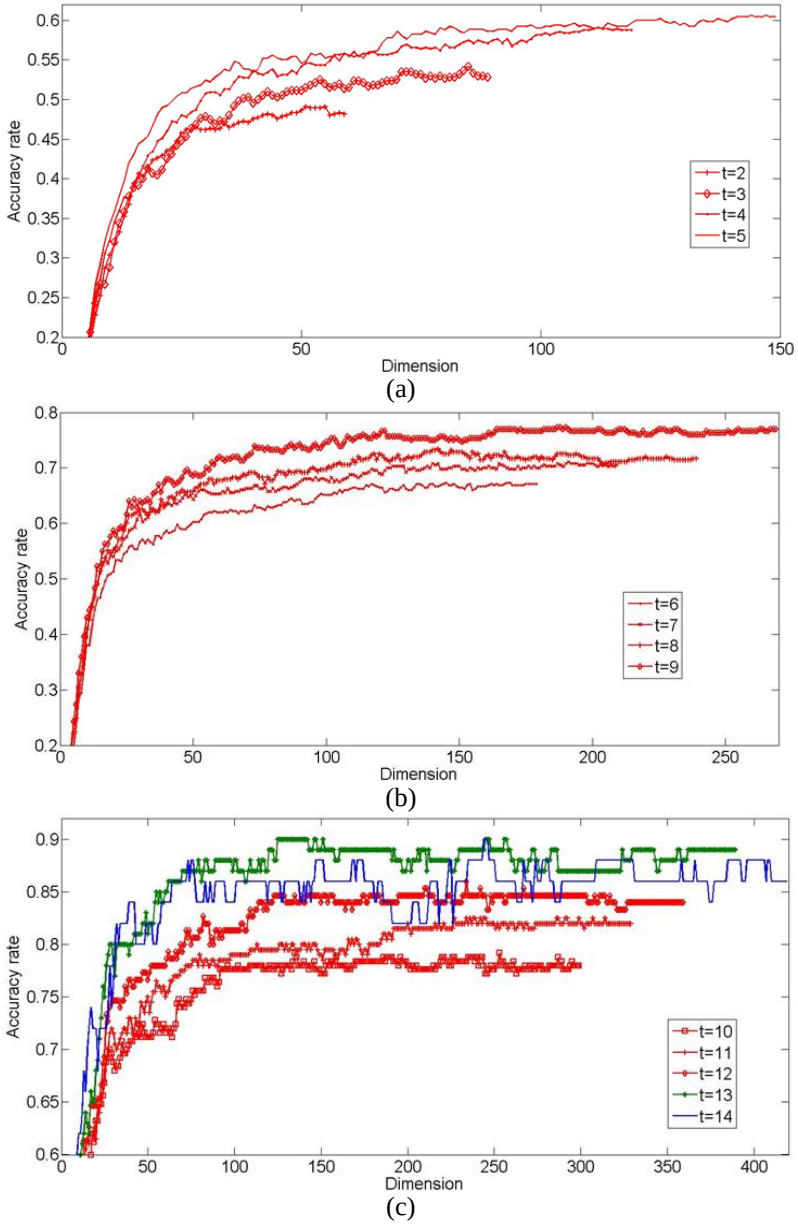


Fig. 6. The accuracies in different dimension in GT face database. (a), (b) and (c) are the results with different number of training samples.

Table 6. The highest accuracy and its value of dimensionality in GT face database

Item	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$	$t=13$	$t=14$
Accuracy	0.4908	0.5417	0.5909	0.6060	0.6733	0.7100	0.7343	0.7733	0.7920	0.8250	0.8600	0.9000	0.9000
Dimensionality	51	85	112	144	145	191	142	187	135	227	234	125	244

Fig. 6 and **Table 6** show the fact that the accuracies have a big improvement according to the increment **o**f dimensionality when the dimensionality is small. After that, the growth rates are smooth when the dimensionality is large enough. But the highest accuracy always does not happen in the largest dimensionality.

4.3 Experiments on UMIST face database

The UMIST face database consists of 564 images of 20 individuals. Each individual is shown in ranges of postures from profile to frontal views. In our experiments, we select the first 19 face images each individual, total 380 face images, to form the subset. **Fig. 7** shows some face images in our experiments.

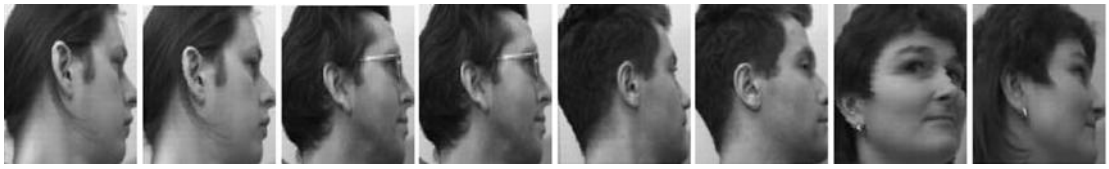


Fig. 7. Some face images from UMIST database in our experiments.

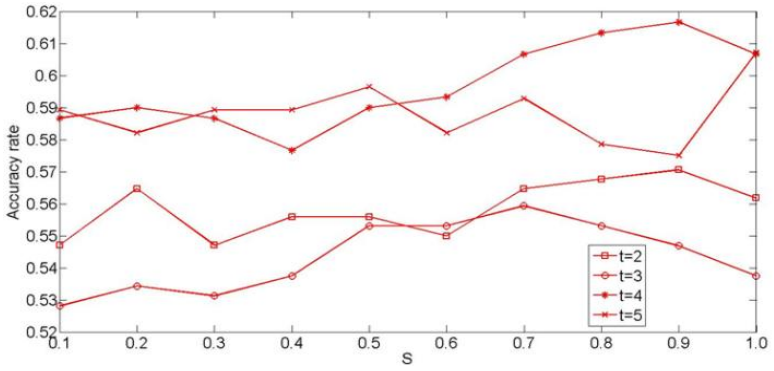
Similar with the above experiment setting, we select the first t $(t = 2, 3, \cdots, 12)$ of each individual for training set and the remaining samples for test. **Table 7** gives the comparative results of different methods in the UMIST face database.

Table 7. Comparative result of different methods on UMIST face database

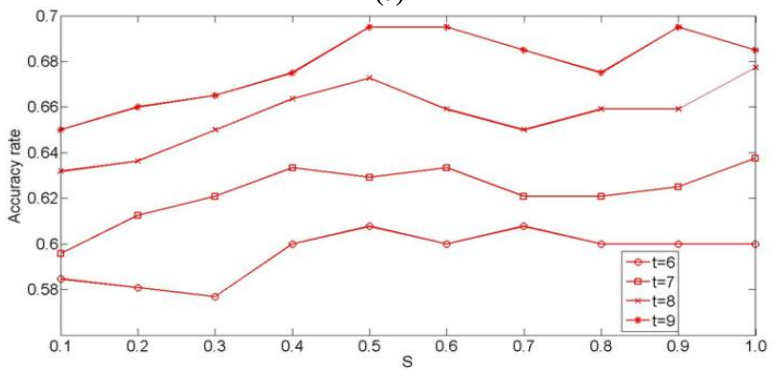
Method	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$
SRC	0.4765	0.5031	0.4967	0.5214	0.5538	0.5875	0.6455	0.6800	0.6889	0.7000	0.7214
CRC	0.4765	0.5031	0.4967	0.5214	0.5538	0.5875	0.6455	0.6800	0.6889	0.7000	0.7214
KCRC	0.5500	**0.5781**	0.5733	**0.6214**	**0.6423**	0.6667	0.6864	**0.7300**	0.7111	0.7375	0.7714
TPTSR	0.2706	0.2375	0.2467	0.2168	0.2500	0.3038	0.3318	0.3000	0.4278	0.4125	0.4429
the proposed method	**0.5588**	0.5625	**0.6067**	0.6000	0.6269	**0.6818**	**0.7150**	0.7000	**0.7438**	**0.7838**	**0.7857**

From **Table 7** we can find that KCRC and our method that are based on kernel space get the highest performance. But TPTSR get the much less accuracies than any other methods. It seems that the large posture variations of the UMIST face database leads to the lack of training samples problem for TPTSR. Though it has a similar tactics to get the sparse representation under ℓ_2 regularization, our method has much higher improvement than TPTSR. Maybe it is more separable for large posture variations in kernel space. Compared with KCRC, the proposed method is able to get a higher performance except $t = 3,5,9$. Though the proposed method cannot get higher accuracies than KCRC in any cases, the average accuracy is higher than that of KCRC. In general, the proposed method is effective and robust.

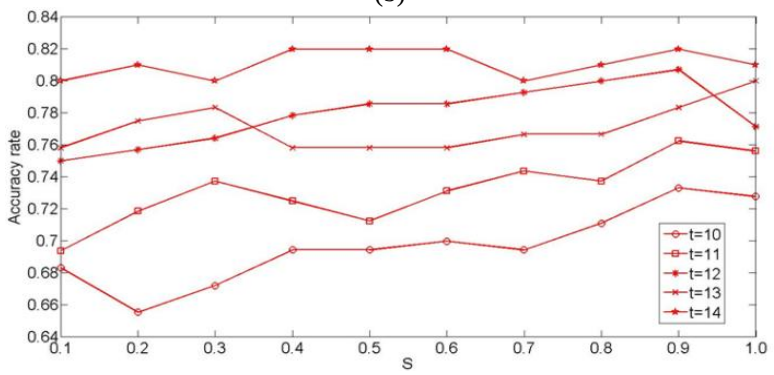
Similar with Section 4.1 and Section 4.2, we investigate the relationship between accuracy and sparseness coefficient, and the relationship between accuracy and dimensionality. The experiment results are shown in **Fig. 8**, **Table 8** and **Fig. 9**, **Table 9**.



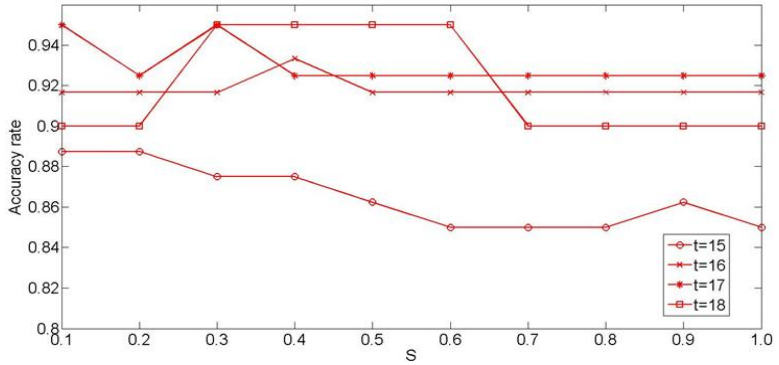
(a)



(b)



(c)



(d)

Fig. 8. The accuracies in different sparseness coefficient S (from 0.1 to 1.0) in UMIST face database. (a), (b), (c) and (d) are the result with different number of training sample.

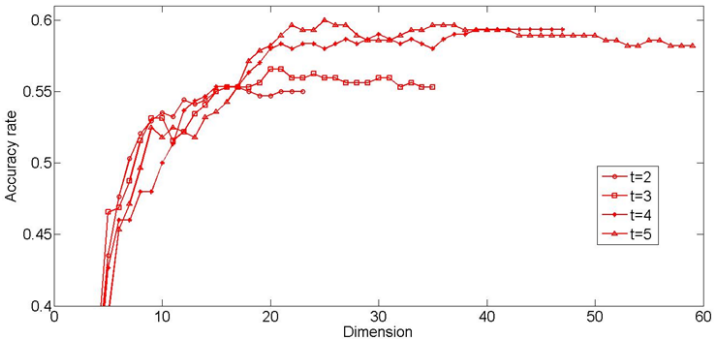
From **Fig. 8** we can find something different compared with Section 4.1 and Section 4.2. At most situation, the high accuracies need more training samples. In another words, the accuracies will be higher when s is bigger in UMIST face database.

Table 8. The highest accuracy and its value of sparseness coefficient (s) in UMIST face database.

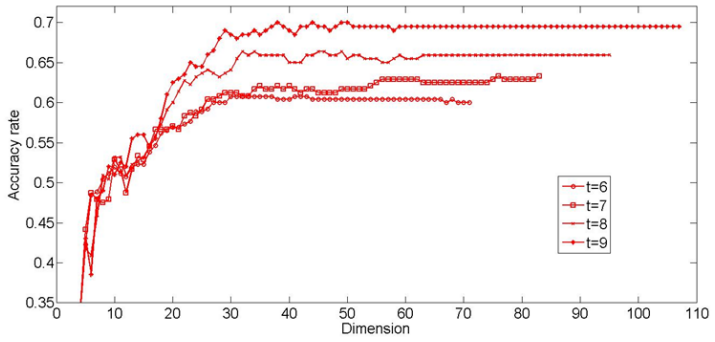
Item	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$
Accuracy	0.5706	0.5594	0.6167	0.6071	0.6077	0.6375	0.6773	0.6950	0.7333	0.7625	0.8071
s	0.1	0.7	0.9	1.0	0.5	1.0	1.0	0.5	0.9	0.9	0.9

From **Table 8** we can see that we can get the highest accuracies when s is enough large except the case when $t=2$. From **Fig. 7** we can find that the face images from UMIST have big varieties in ranges of postures from profile to frontal views. Maybe these big varieties aggravate the SSS problem. So it needs more training samples to represent the test samples well.

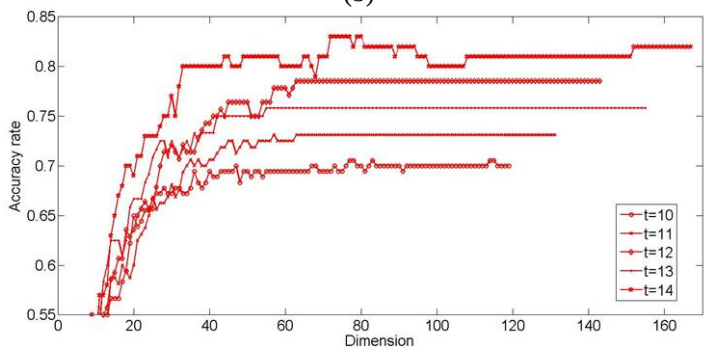
Fig. 9 and **Table 9** show the relationship between accuracy and dimensionality.



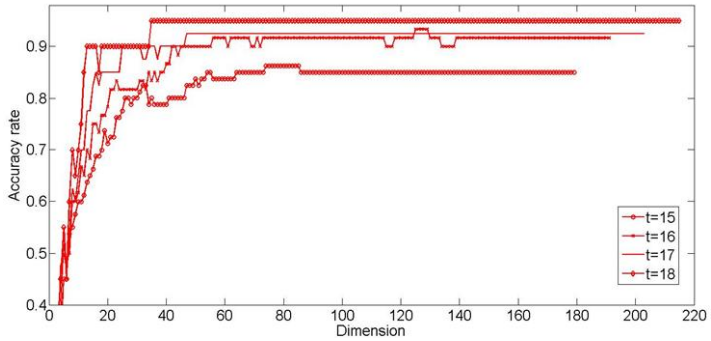
(a)



(b)



(c)



(d)

Fig. 9. The accuracies in different dimension in the UMIST face database. (a), (b), (c) and (d) are the results with different number of training samples.

Fig. 9 shows that the accuracy improves more quickly when dimension is small. When the dimensionality is large enough the accuracy seems stable. We also can find that the highest accuracy does not appear in the highest dimension in any case of t. In another words, we do not need all the information of the training samples for classification though face recognition is a classical SSS problem.

Table 9. The highest accuracy and the dimension in UMIST face database.

Item	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$
Accuracy	0.5529	0.5656	0.5933	0.6000	0.6077	0.6333	0.6636	0.7000	0.7056	0.7312	0.7857
Dimensionality	16	20	39	25	30	76	32	38	77	57	63

From **Table 9** we can find that we need much smaller dimensionality than that in the ORL and the GT face databases to get the highest accuracy.

5. Conclusions

In this paper, an improved kernel principal component analysis method based on spare representation is proposed for more performance in face recognition. Our method implements sparse representation under ℓ_2 regularization based on collaborative representation in kernel space. In order to improve the robustness, the proposed method extracts the most relevant features for test sample twice. The comparative experiments are conducted in the ORL, the GT and the UMIST face databases. Also, we investigate the relationship between accuracy and sparseness coefficient by fixing the dimensionality, the relationship between accuracy and dimensionality by fixing the sparseness coefficient respectively. The compared methods include several stat-of-the-art face recognition methods including SRC, CRC, KCRC and TPTSR. The experimental results verify the efficiency and robustness of the proposed method. From the expreimental results, we are able to observe that the sparseness coefficient and dimensionality are all important for accuracy. In the future works, we will try to find out the concrete relationship of the two parameters and improve the accuracy.

Acknowledgements

This work is partially supported by Specialty Construction Project in Guangdong Province (No.2013LYM 0055), the Scientific Project of Hanshan Normal University (LY201301) and the Science and Technology Program of Chaozhou City (2012S13), the High Level Talents in Colleges of Guangdong Province (Guangdong Finance Education [2013] No. 246), and the Natural Science Foundation of Guangdong Province (No. 2014A030307014).

Reference

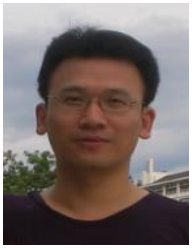
[1] W. Zhao, R. Chellappa, A. Rosenfeld and P. Phillips, "Face recognition: a literature survey," *ACM Computing Serveys*, vol. 3, no. 4, pp. 399-458, 2003. Article (CrossRef Link).
[2] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expression: A survey," *Pattern Recognition*, vol. 25, no.1, pp. 65-77, 1999. Article (CrossRef Link).
[3] W. Huang and H. Yin, "On nonlinear dimensionality reduction for face recognition," *Image and Vision Computing*, vol. 30, no.4-5, pp. 355-366, 2012. Article (CrossRef Link).
[4] P.C. Hsieh and P.C. Tung, "Shadow compensation based on facial symmetry and image average for robust face recognition," *Neurocomputing*, vol. 73, no. 13-15, pp. 2708-2717, 2010. Article (CrossRef Link).
[5] Y. Xu, X. Fang, X. Li, J. Yang, J. You, H. Liu and S. Teng, "Data uncertainty in face recognition," *IEEE Transactions on Cybernetics*, vol. 44, no. 10, pp. 1950-1961, 2014. Article (CrossRef Link).
[6] J. Lu, K.N. Plataniotis and A.N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 3079-3087, 2003. Article (CrossRef Link).
[7] Y. Xu, X. Li, J. Yang and D. Zhang, "Integrate the original face image and its mirror image for face recognition," *Neurocomputing*, vol. 131, pp. 191-199, 2014. Article (CrossRef Link).
[8] W. Huang, X. Wang, Y. Ma, Y. Jiang, Y. Zhu and Z. Jin, "Robust Kernel Collaborative Representation for Face Recognition," *Optical Engineering*, vol. 54, no.5, pp. 053103-1-053103-10, 2015. Article (CrossRef Link).
[9] Y. Xu, X. Li, J. Yang, Z. Lai and D. Zhang, "Integrating conventional and inverse representation for face recognition," *IEEE Transactions on Cybernetics*, vol. 44, no. 10, pp. 1738-1746, 2014. Article (CrossRef Link).
[10] M.Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol.3, no. 1, pp. 71-86, 1991. Article (CrossRef Link).
[11] D.L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.18, no. 8, pp. 831-836, 1996. Article (CrossRef Link).
[12] K.R. Muller, "An introduction to kernel-based learning algorithms," *IEEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181-201, 2001. Article (CrossRef Link).
[13] Y. Jian, Z. David, A.F. Frangi and Y. Jing-Yu, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no.1, pp. 131-137, 2004. Article (CrossRef Link).
[14] J. Yang, D. Zhang, X. Yong and J.Y. Yang, "Two-dimensional discriminant transform for face recognition," *Pattern Recognition*, vol. 38, no. 7, pp. 1125-1129, 2005. Article (CrossRef Link).
[15] Z. Jin, J.Y. Yang, Z.S. Hu and Z. Lou, "Face recognition based on uncorrelated discriminant transformation," *Pattern Recognition*, vol. 34, no. 7, pp. 1405-1416, 2001. Article (CrossRef Link).
[16] A.M. Jadea, B. Srikanth, V.K. Jayaraman, B.D. Kulkarni, J.P. Jog and L. Priya, "Feature extraction and denoising using kernel PCA," *Chemical Engineering Science*, vol. 58, no.19, pp. 4441-4448, 2003. Article (CrossRef Link).

[17] H. Hoffmann, "Kernel PCA for novelty detection," *Pattern Recognition*, vol. 40, no. 3, pp. 863-874, 2007. Article (CrossRef Link).
[18] F. Ye, Z. Shi and Z. Shi, "A Comparative Study of PCA, LDA and Kernel LDA for Image Classification," in: *ISUVR '09 International Symposium on*, pp. 51-54, 2009. Article (CrossRef Link).
[19] X. Fang, Y. Lu, Z. Li, L. Yu and Y. Chen, "Kernel representation-based nearest neighbor classifier," *Optik*, vol. 125, pp. 2320-2326, 2014. Article (CrossRef Link).
[20] C. Lu, C. Zhang, T. Zhang and W. Zhuang, "Kernel based symmetrical principal component analysis for face classification," *Neurocomputing*, vol. 70, pp. 904-91, 2007. Article (CrossRef Link).
[21] G. Heo, P. Gader and H. Frigui, "RKF-PCA: Robust kernel fuzzy PCA," *Neural Networks*, vol. 22, pp. 642-650, 2009. Article (CrossRef Link).
[22] R. Zhang, W. Wang and Y. Ma, "Approximations of the standard principal components analysis and kernel PCA," *Expert Systems with Applications*, vol. 37, pp. 6531-6537, 2010. Article (CrossRef Link).
[23] Y.W. Zhang, H. Zhou and S.J. Qin, "Decentralized Fault Diagnosis of Large-scale Processes Using Multiblock Kernel Principal Component Analysis," *Acta Automatica Sinica*, vol. 36, no. 4, pp. 593-597, 2010. Article (CrossRef Link).
[24] S.W. Choi, C. Lee, J.M. Lee, J.H. Park and I. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, pp. 55-67, 2005. Article (CrossRef Link).
[25] T.J. Hansen, T.J. Abrahamsen and L.k. Hansen, "Denoising by semi-supervised kernel PCA preimaging, " *Pattern Recognition Letter*, vol. 49, pp. 114-120, 2014. Article (CrossRef Link).
[26] A.E. Mercer, M.B. Richman and L.M. Leslie, "Identification of severe weather outbreaks using kernel principal component analysis," *Procedia Computer Science*, vol. 6, pp. 231-236, 2011. Article (CrossRef Link).
[27] H. Sahbi, "Kernel PCA for similarity invariant shape recognition," *Neurocomputing*, vol. 70, pp. 3034-3045, 2007. Article (CrossRef Link).
[28] X. Li and L. Shu, "Kernel based nonlinear dimensionality reduction for microarray gene expression data analysis," *Expert Systems with Applications*, vol. 36, pp. 7644-7650, 2009. Article (CrossRef Link).
[29] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 210–227, 2009. Article (CrossRef Link).
[30] L. Zhang, M. Yang and X. Feng, "Sparse Representation or Collaborative Representation: Which Helps Face Recognition?" in *Proc. of IEEE International Conference on Computer Vision*, pp. 471-478, 2011. Article (CrossRef Link).
[31] C.X. Ren, D.Q. Dai and H. Yan, "Robust classification using $\ell_{2,1}$ -norm based regression model," *Pattern Recognition*, vol. 45, pp. 2708-2718, 2012. Article (CrossRef Link).
[32] Y. Xu, Q. Zhu, Z. Fan, D. Zhang, D. Zhang, J. Mi and Z. Lai, "Using the idea of the sparse representation to perform coarse to fine face recognition," *Information Sciences*, vol. 238, pp. 138-148, 2013. Article (CrossRef Link).
[33] P. Zhu, L. Zhang, Q. Hu, and S.C.K. Shiu, "Multi-scale Patch based Collaborative Representation for Face Recognition with Margin Distribution Optimization," *Lecture Notes in Computer Science*, vol. 7572, pp. 822-835, 2012. Article (CrossRef Link).
[34] Y. Xu, D. Zhang, J. Yang and J.Y. Yang, "A two-phase test sample sparse representation method for use with face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 9, pp. 1255-1262, 2011. Article (CrossRef Link).
[35] F. Dornaika, Y. Traboulsi and A. Assoum, "Adaptive two phase sparse representation classifier for face recognition," *Advanced Concepts for Intelligent Vision Systems*, pp. 182-191, 2013. Article (CrossRef Link).

[36] X. Shi, Y. Yang, Z. and Z. Lai, "Face recognition by sparse discriminant analysis via joint $L_{2,1}$ -norm minimization," *Pattern Recognition*, vol. 47, no. 7, pp. 244-2453, 2014. Article (CrossRef Link).
[37] C. Hou, F. Nie, D. Yi and Y. Wu, "Feature Selection via Joint Embedding Learning and Sparse Regression," in *Proc. of the Twenty-Second International Joint Conference on Artificial Intelligence*, pp. 1324-1329, 2012. Article (CrossRef Link).
[38] S. Ghosh, "On the grouped selection and model complexity of the adaptive elastic net," *Statistics and Computing*, vol. 21, no. 3, pp. 451-462, 2011. Article (CrossRef Link).
[39] Y. Xu, W. Zuo and Z. Fan, "Supervised sparse representation method with a heuristic strategy and face recognition experiments," *Neurocomputing*, vol. 79, no. 1, pp. 125-131, 2012. Article (CrossRef Link).
[40] J. Jiang, H. Gan, L. Jiang, C. Gao and N. Sang, "Semi-supervised Discriminant Analysis and Sparse Representation-based self-training for Face Recognition," *Optik*, vol. 125, no. 9, pp. 2170-2174, 2014. Article (CrossRef Link).
[41] A. Eftekhari, M. Forouzandfar, H. A. Moghaddam and J. Alirezaie, "Block-wise 2D kernel PCA/LDA for face recognition," *Information Processing Letters*, vol. 110, pp. 761-766, 2010. Article (CrossRef Link).
[42] Y. Xu, X. Zhu, Z. Li, G. Liu, Y. Lu and H. Liu, "Using the original and 'symmetrical face' training samples to perform representation based two-step face recognition," *Pattern Recognition*, vol. 46, no. 4, pp. 1151-1158, 2013. Article (CrossRef Link).
[43] J. Wang, H. Bensmail and X. Gao, "Feature selection and multi-kernel learning for sparse representation on a manifold," *Neural Networks*, vol. 51, pp. 9-16, 2014. Article (CrossRef Link).
[44] R.G. Baraniuk, "Compressive Sensing [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp.118-121, 2007. Article (CrossRef Link).
[45] J. Zhang, C. Zhao, D. Zhao and W. Gao, "Image compressive sensing recovery using adaptively learned sparsifying basis via L_0 minimization," *Signal Processing*, vol. 103, pp. 114-126, 2014. Article (CrossRef Link).
[46] G. Shenghua, T.W. Hung and C.L. Tien, "Sparse Representation with Kernels," *IEEE Transactions on Image Processing*, vol. 22, no 2, pp. 423-434, 2013. Article (CrossRef Link).
[47] W. Yang, Z. Wang, J. Yin, C. Sun and K. Ricanek, "Image classification using kernel collaborative representation with regularized least square," *Applied Mathematics and Computation*, vol. 222, pp.13-28, 2013. Article (CrossRef Link).
[48] http://www.cam-orl.co.uk
[49] http://www.anefian.com/face_reco.htm
[50] http://www.sheffield.ac.uk/eee/research/iel/research/face
[51] H.K. Palo and M.N. Mohanty, "Classification of Emotions of Angry and Disgust," *Smart Computing Review*, vol. 5, no. 3, pp.151-158, 2015. Article (CrossRef Link).

Wei Huang is now an associate professor with the School of Computer and Information Engineering, Hanshan Normal University, China. He received his M.S. degree in computer application technology in 2008 from Nanjing University of Science and Technology (NUST), and his B.S. degree in Hanshan Normal University in 2001. His research interest includes pattern recognition, machine learning and image processing.

Wang Xiaohui received her B.S. degree in mechanics engineering from the Hunan University, Changsha, China, in 2000, and the M.S. Degree in Computer Science from the Harbin institute of technology, Harbin, China, in 2006. From 2000 to 2003, she was an assistant at the department of Physics in the south of Fujian Normal University. From 2006 to now, she serves as lecturer at School of Computer and Information Engineering in Hanshan Normal University. Her research interests include image processing, pattern recognition and statistical signal processing.



Yinghui Zhu received the B.S. degree in computer science from South China Normal University, Guangzhou, China, in 2005, and the M.S. degree in computer science from South China Normal University, Guangzhou, China, in 1999. He serves as associate professor at School of Computer and Information Engineering in Hanshan Normal University, China. His research interests include graphic and image processing, artificial intelligence.

Gengzhong Zheng was born in 1975. He received his Ph.D. degree in School of Computer Science and Technology from Xidian University in 2012. He is working as a professor at School of Computer and Information Engineering in Hanshan Normal University. His current research interests include network optimization and wireless sensor networks.