

# Single-sample face recognition based on improved SRC and expanding sample

Zhijing Xu<sup>a</sup>, Li Ye<sup>a</sup>, Xiangjian He<sup>b</sup>

<sup>a</sup> Shanghai Maritime University, Shanghai, China

<sup>b</sup> University of Technology Sydney, Australia

## Abstract

This paper proposes a kind of face recognition method with one training image per person, which is based on compressed sensing. There are two methods—nonlinear dimensionality reduction by locally linear embedding and sparse coefficients, by which redundant samples can generate. These new samples with multi-expressive and multi-gesture can be treated as training samples. Finally a super sparse random projection and weighted optimization have been used to improve the SRC (weighted super sparse representation classification, WSSRC). The improved SRC algorithm can be applied to face recognition. Experiments on the well-known ORL face database and FERET face database show that the presented method is respectively about 15.53% and 7.67%, more accurate than original SRC method in the context of single sample face recognition problem. In addition, extensive experimentation reported in this paper suggests that the presented method achieve higher recognition rate than RSRC, SSRC DMMA, and DCT-based DMMA.

## 1. Introduction

Face recognition technology is a computer technology for identity authentication by analyzing and comparing facial feature information <sup>[1]</sup>. Because the two-dimensional image of face is greatly impacted by light face, facial expressions, gestures, face recognition is still an unresolved problem. At the same time, due to difficulties in collecting the sample and the limited storage capacity of the system cause the single-sample problem, which makes the extraction of intrinsic features of training objects become more difficult <sup>[2]</sup>. In the case that per person has only one image in the face training database, the recognition performance of the majority of the conventional methods will be severely degraded. Compressed Sensing <sup>[3]</sup> (CS) is an emerging signal processing methods in recent years. By use of compressed sensing theory, Wright et al. presented a sparse representation classification algorithm (SRC) <sup>[3]</sup>. SRC method can effectively respond to changes of lighting, facial expression and partial occlusions, so it has a good robustness. In order to improve the recognition rate, we use a super sparse random projection and weighted optimization to improve the SRC. For single-sample problem, we propose a single-sample face recognition method based on the improved SRC. Firstly, it makes a face image become pieces by use of local neighborhood embedding nonlinear dimension reduction and sparse coefficients, and then regards these generated new multi-expression and multi-gesture samples as training samples, and finally uses the improved SRC to identify and classify.

## 2. CS and WSSRC

### 2.1 CS

Compressed sensing presented by Donoho and Candes is a new theoretical framework. Its main model is linear. The basic idea is that high-dimensional original signal can be effectively restored by taking advantage of the low-dimensional observation signal as long as the signal meets the "sparse" feature or the signal is compressible.

Compressed sensing theory mainly covers the following aspects:

(1) For the signal  $\mathbf{x} \in \mathbf{R}^n$ , how to find an orthogonal basis or tight frame  $\Psi$  to make it can be sparse in  $\Psi$ , that is to say, the sparse representation problem of a signal:  $\mathbf{x} = \Psi\alpha$ .

(2) how to design a stable  $M \times N$ -dimensional measurement matrix  $\Phi$  unrelated to  $\Psi$ , which can ensure the important information of the signal is not damaged from the  $N$ -dimension to  $M$ -dimensional, that is to say, the low-speed signal sampling problem:  $y = \Phi x$ .

(3) How to design fast reconstruction algorithm to from recover signal from linear observation  $y = A\alpha$  ( $A$  is called CS information operator,  $A = \Phi\Psi$ ), that I to say, the reconstruction problems of the signal. It is based on the rigorous mathematical optimization question:

$$\min_{\alpha} \|\hat{\alpha}\|_0 \quad \text{s.t.} \quad A\hat{\alpha} = y \quad (1)$$

L0 norm optimization problem is a NP problem, that is to say, it is difficult to solve in polynomial time, or even impossible to verify the reliability of the solution of the problem. So, we have to change L0 norm to L1 norm:

$$\min_{\alpha} \|\hat{\alpha}\|_1 \quad \text{s.t.} \quad A\hat{\alpha} = y \quad (2)$$

## 2.2. SRC

The basic idea of SRC is: the original data set which contains the all training samples can be seen as over-complete dictionary, so the test samples can be characterized sparsely by the over-complete dictionary. If each class has enough training samples, the test samples can only be represented as a linear combination of the training samples from the same class. For the over-complete dictionary, it is the best sparse representation of test samples, and can be obtained by the L1 norm minimization. Therefore, finding the best sparse representation can naturally distinguish different types of training samples.

On the face recognition problem, if there are enough training samples in class  $i$ :  $A_i = [v_{i,1}, v_{i,2}, \dots, v_{i,N_i}]$ , a new face image (test sample)  $x \in \mathbf{R}^n$  can be represented as a linear combination of these given training samples:  $x = A_i \alpha_i$ ,  $\alpha_i \in \mathbf{R}^{N_i}$  is the coefficient vector. If the test sample's category is unknown, cascade all  $C$  categories of training samples whose number is  $N = \sum_{i=1}^C N_i$  together to form a new matrix  $\Psi$  to represent the entire training set:

$$\Psi = [A_1, A_2, \dots, A_C] = [v_{1,1}, v_{1,2}, \dots, v_{C,N_C}]$$

So the  $x$  can be expressed linearly by all training samples:

$$x = \Psi \alpha \in \mathbf{R}^n \quad (3)$$

Ideally  $\alpha = [0, 0, \dots, 0, \alpha_i^T, 0, 0, \dots, 0]^T$  is a coefficient vector. The vector's other elements are 0 in addition to those elements related to class  $i$ .

In actual recognition scenarios, the test sample  $x$  may be partially broken or occluded. So the formula (3) can be amended as follows:

$$x = x_0 + e_0 = \Psi \alpha + e_0 \quad (4)$$

In it,  $e_0 \in \mathbf{R}^n$  is an error vector, due to the number of unknowns exceeds the number of equations in the formula (4), it is not direct to solve  $\alpha$ . However, under a certain condition, the desired solution  $(\alpha, e_0)$  is the best sparse solution of the formula (4):

$$\begin{aligned} (\alpha, e_0) &= \arg \min \|\alpha\|_0 \\ \text{s.t. } x &= \Psi \alpha + e_0 \end{aligned}$$

In it, L0 norm refers to the number of non-zero elements in a vector. The literature [3] found  $\alpha$  by solving the following convex question because of the theory of L0 and L1 minimization equivalence:

$$\begin{aligned} \min & \|\alpha\|_1 + \|e\|_1 \\ \text{s.t. } x &= \Psi \alpha + e_0 \end{aligned} \quad (5)$$

Once the L1 minimization problem is solved, use the residual  $r_i = \|x - A_i \alpha_i\|$  to decide the category of the test sample.

### 2. 3. WSSRC

This paper presents the algorithm by using a super sparse random projection matrix as the measurement matrix  $\Phi$  to reduce the image dimension. The literature [4] has proved that the super sparse random projection matrix not only meets the necessary condition of CS measurement matrix but also has better measurement result than the Gaussian measurement matrix. Find the optimal sparse solution by use of dual augmented Lagrange multiplier (DALM) on L1 minimization problem. Its principle is to translate the formula (1) into the dual problem:

$$\max_t \mathbf{y}^T \mathbf{t} \quad \text{s.t.} \quad \mathbf{A}^T \mathbf{t} \in \mathbf{B}_1^\infty$$

In it,  $\mathbf{B}_1^\infty = \{\boldsymbol{\alpha} \in \mathbf{R}^n : \|\boldsymbol{\alpha}\|_\infty \leq 1\}$  The dual problem's corresponding Lagrange function is:

$$\min_{\mathbf{t}, \mathbf{z}} -\mathbf{y}^T \mathbf{t} - \boldsymbol{\alpha}^T (\mathbf{z} - \mathbf{A}^T \mathbf{t}) + \frac{\beta}{2} \|\mathbf{z} - \mathbf{A}^T \mathbf{t}\|_2^2 \quad \text{s.t.} \quad \mathbf{z} \in \mathbf{B}_1^\infty \quad (6)$$

In it,  $\boldsymbol{\alpha}$  is the Lagrange multiplier. Because it is impossible to change  $\boldsymbol{\alpha}$ ,  $\mathbf{t}$ ,  $\mathbf{z}$  at the same time, we can use the alternating iterative method to find the extreme value of the formula(6), that is to say, update the third variable continuously by minimizing the cost function while maintaining the two other variables held constant. This will not only ensure that the reconstruction accuracy but also reduce the algorithm's calculating velocity. Finally, use weighted residual to determine the category of test data for more precise classification.

Algorithm is as follows:

(1) Input the N training samples which have C classes to constitute the dictionary matrix  $\Psi = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_C] \in \mathbf{R}^{n \times N}$ .

(2) Generate column-normalized very sparse random projection matrix  $\Phi \in \mathbf{R}^{M \times n}$ , in it,  $\Phi(i, j) = \frac{1}{\sqrt{M}} \mathbf{h}_{ij}$ , while  $\mathbf{h}_{ij}$  obeys very sparse projection distribution:

$$\mathbf{h}_{ij} \sim \begin{pmatrix} \frac{\sqrt[4]{n}}{2\sqrt{n}} & 0 & -\frac{\sqrt[4]{n}}{2\sqrt{n}} \\ \frac{1}{2\sqrt{n}} & 1 - \frac{1}{\sqrt{n}} & \frac{1}{2\sqrt{n}} \end{pmatrix}$$

(3) Give a test image  $\mathbf{x} \in \mathbf{R}^n$ .

(4) Calculate  $\mathbf{Y}_i = \Phi \mathbf{A}_i$ , ( $i = 1, \dots, C$ ), and calculate the measurement matrix  $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_C] = \Phi \Psi$ , in it  $\mathbf{Y} \in \mathbf{R}^{M \times N}$ .

(5) Calculate the new projection sample  $\mathbf{y} = \Phi \mathbf{x}$ , and find the sparse vector  $\hat{\boldsymbol{\alpha}}$  by using DALM to make  $\mathbf{Y} \boldsymbol{\alpha} = \mathbf{y}$ .

(6) Let  $\hat{\boldsymbol{\alpha}} = [\mathbf{r}_1^T, \mathbf{r}_2^T, \dots, \mathbf{r}_N^T]$ , calculate the reconstruction error by use of weighted residual  $\mathbf{e}_i = \frac{\|\mathbf{y} - \mathbf{Y}_i \mathbf{r}_i\|}{\|\mathbf{r}_i\|}$  in each class .

(7) If  $k = \arg \min_i (\mathbf{e}_i)$ , then  $\mathbf{x}$  belongs to the class k.

### 3. Generation of redundant sample

The difficulty of single sample problem is that each object has only a sample. Adding redundant information of samples is an effective way to solve the single sample problem. This article firstly single training sample use the method of local neighborhood embedding nonlinear dimension reduction<sup>[6]</sup> to make single training sample become pose-varied images. Then generate multi-expression images by use of sparse coefficient. Finally, make these multi-expression and multi-gesture new samples as training samples for training. Its purpose is to get more useful information from only one training sample in order to identify further classification.

#### 3. 1. Generation of multi-gesture samples

The method of generating multi-gesture samples is based on the theory of local neighborhood embedding nonlinear dimension reduction. The specific implementation is:

(1) Let all face images represent in form of a column vector.

(2) Set the input face image which has one gesture as  $I_i$ . Regard  $I_i$  as one point of high-dimensional space, and the same gesture training set of face images  $T_i^n$  (including N images) is regarded as near point of  $I_i$ . Then find the weight value of near points by local neighborhood embedding nonlinear dimension reduction.

(3) Set the target gesture training set of face images as  $T_o^n$  (including N images), and the unknown face image which has the target gesture is  $I_o$ . Then use the weight value solved in the step (2) to calculate a point of that high-dimensional space, that is to say, find the  $I_o$ .

(4) Let the target pose face image vector represent in form of matrix.

This method can be easily and quickly to generate multi-pose samples, and can overcome the problem of complexity, large amount of computation, and non-ideal result of those existing similar algorithms. Figure 1 is the original face images and generated new samples.

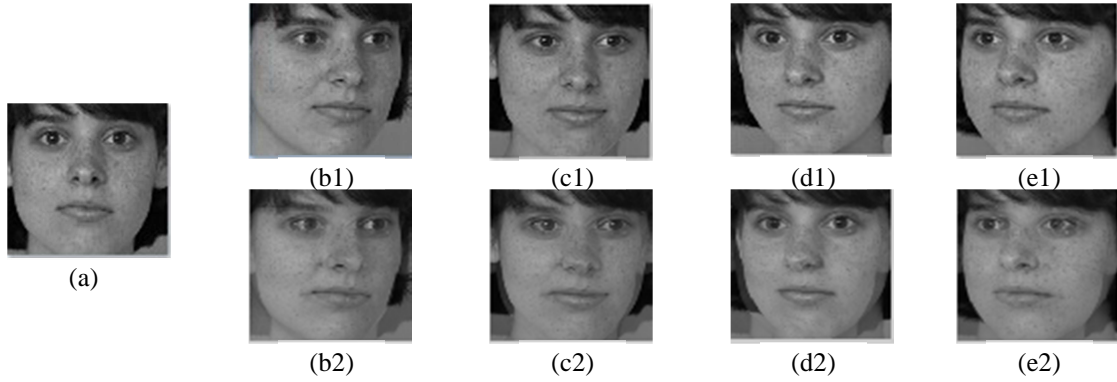


Figure1 the original face images and generated new samples

### 3. 2. Generation of multi-expression samples

The method of generating multi-expression samples is based on sparse representation. Make sure that all images' size uniform and the eye, nose, mouth of every image are fixed in the same position of them. For a single frame face test image which has a kind of expressive, use a linear combination of face images having the same expression to approximate the test image, so we can get the sparse reconstruction coefficients of the approaching linear combination. Then take advantage of the sparse reconstruction coefficients and target expression training set of face images to reconstruct the face image which has the target expression.

This method is simple and does not contain troublesome facial feature extraction step, so it is easy to implement. Besides, its algorithm complexity is relatively reduced. Figure 2 is the original face images and generated multi-expression samples., (a) the input happy face test image; (b1), (c1), (d1), (e1), (f1) respectively is the generated anger, disgust, fear, sadness, surprise face image, and (b2), (c2), (d2), (e2), (f2) is the corresponding real images.

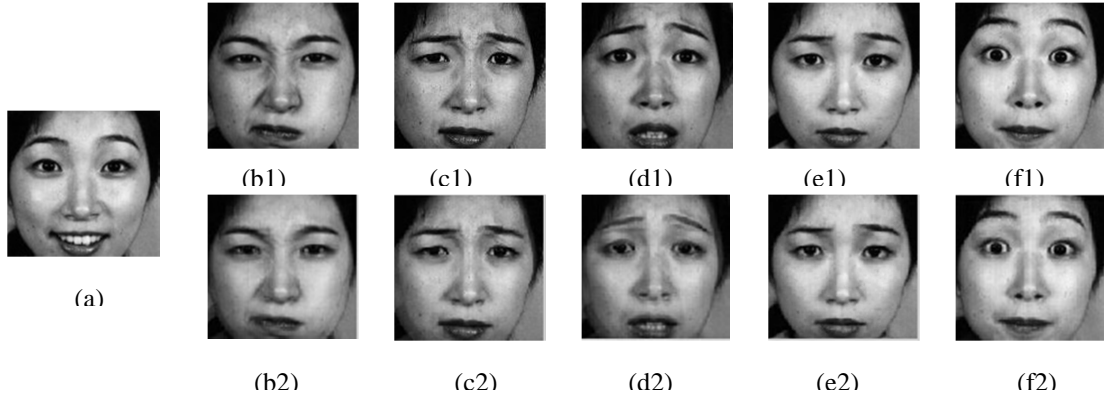


Figure2 the original face images and generated new samples

#### 4. Design of experiment and simulation results

The paper uses ORL, FERET face database for experimental verification. Select a face image for training, and the remaining images are regarded as test samples. Readjust all face images' size as  $32 \times 32$  before the experiment. In the case of a single sample, this paper first takes the presented method to generate multi-expression and multi-pose new samples, which is based on gesture and expression library. Then the original training sample and new samples are regarded as training samples. Finally, use the improved SRC for classification.

##### 4. 1. ORL

ORL database has total 40 individuals' 400 pictures, that is to say, each person has 10 images. Some images are taken at different times, and human facial expressions and facial details have varying degrees of change. In the experiment, take a sample of each person to expand. Then let the generated redundant samples and the original sample together as training samples and samples the remaining images as test samples. Figure 3 depicts the recognition rate while using a different number of redundant samples.

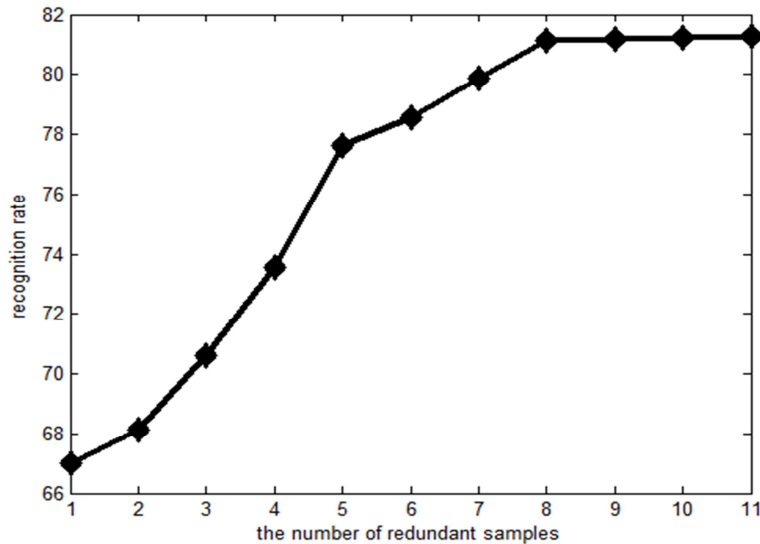
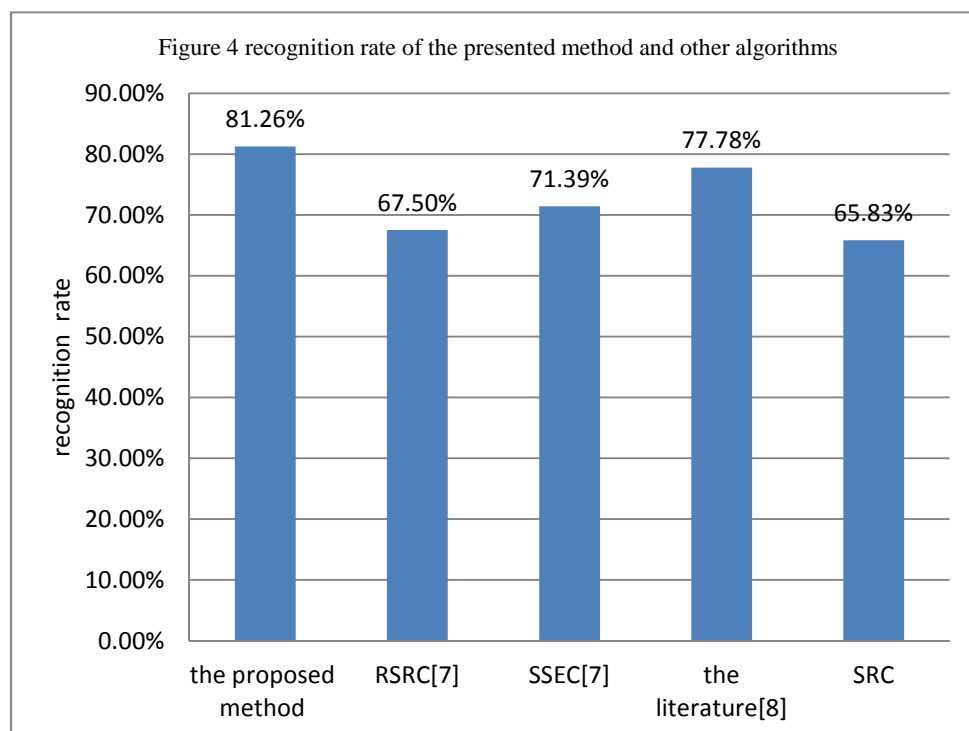


Figure3 recognition rate under different number of redundant sample

As can be seen from Figure3, the recognition rate is low when the number of redundant sample is not enough. With the increase of the number of redundant sample, the recognition rate

significantly increased. But when the redundant sample reaches 8, the recognition rate grows very slowly, almost gently. The reason is that the number of samples is the main cause of influencing the identification effect when the intrinsic characteristics of these samples are too few because of the too few training samples. But with the number of redundant sample continues to improve, it will be bound to cause a lot of redundant information, and the data distribution of the generated redundant samples has some inconsistency with the data distribution of the original samples, which can indicate that increasing the number of redundant sample has a limited influence on the identification rate.

Compare the algorithm of this paper with SRC and the algorithms in references7, 8. As can be seen from Figure 4, the method mentioned in this paper improves 15.53% than the original SRC method, and has better recognition performance than those algorithms in references7, 8.



#### 4.2. FERET

Take some images from FERET face database, including 200 individuals' 1400 pictures. They are taken under conditions that expression, perspective and light intensity are different. Similarly, take a sample of each person to expand, and the remaining images as test samples. Figure 5 depicts the recognition rate while using a different number of redundant samples.

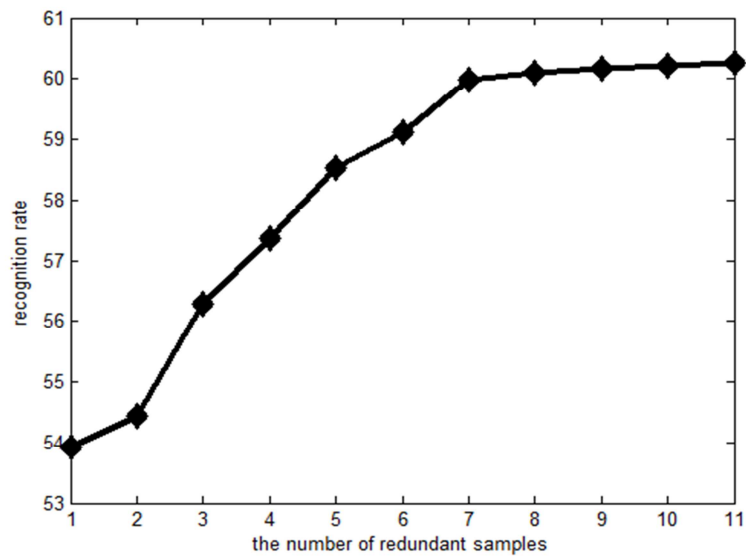
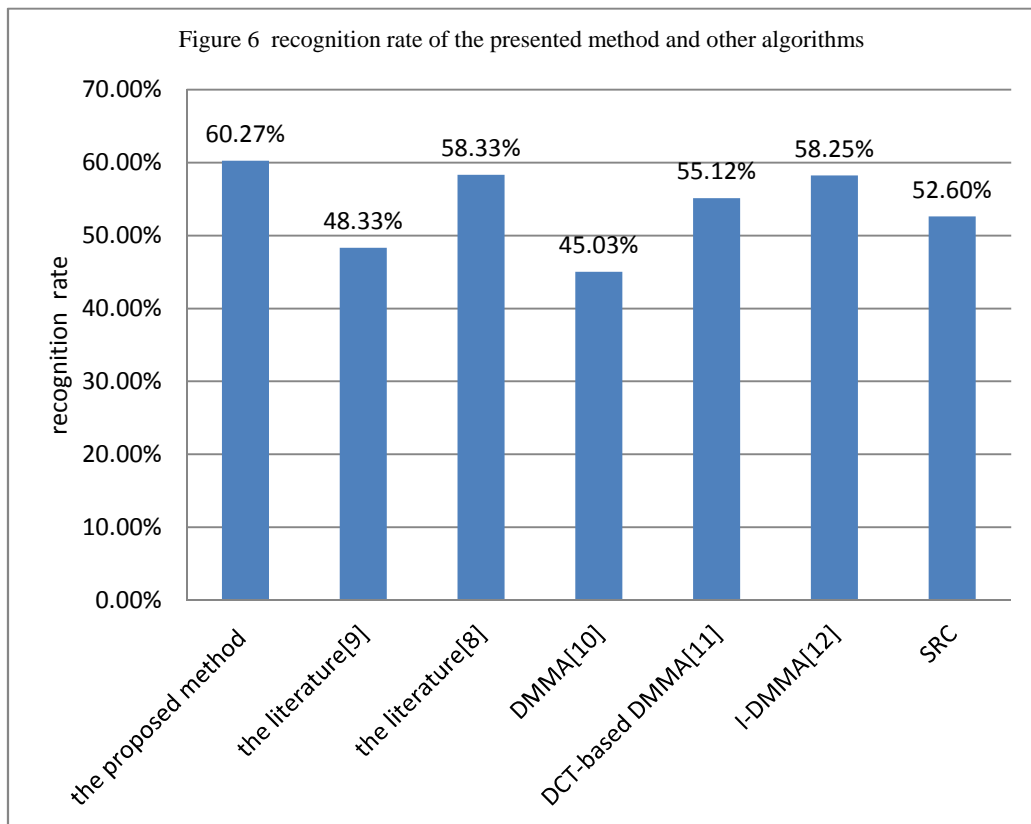


Figure 5 recognition rate under different number of redundant sample

Some researches about single-sample face recognition were also done in literatures [8-12]. It can be seen from Figure 6 that the presented method increases 7.63% than the original SRC on FERET face database. The presented method also has better recognition performance than DMMA, DCT-based DMMA, I-DMMA, etc. Figure 6 shows the single-sample recognition rate of these methods applied on the FERET.



## 5. Conclusion

For single-sample problem of face recognition research, we propose a CS-based single-sample face recognition algorithm. These comparison experiments show that

the method not only takes advantage of the feature information of a single sample to generate some new images, but also greatly improves the recognition rate in the case of single sample, which provides a new method for single-sample face recognition technology.

### Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 61404083 and Foundation of Shanghai Maritime University Grant No. 20120108.

### References

- [1]Mingjun Wang. Study of Face Recognition Algorithm Based on Compressive Sensing[D]. MS Thesis of Xi'an Electronic and Engineering University, 2012.
- [2]Wang Jie, Plataniotis K N, Lu Juwei, et al. On Solving the Face Recognition Problem with One Training Sample per Subject[J].Pattern Recognition, 2006, 39(9): 1746-1762.
- [3]Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009, 31(2): 210-227.
- [4]Hong Fang, Quanbin Zhang, Sui Wei. Method of image reconstruction based on very sparse random projection.ComputerEngineering and Applications, 2007,43(22):25-27.
- [5]Allen Y. Zihan Zhou. Arvind Ganesh Balasubramanian. S. Shankar Sastry. Yi Ma.Yang, Fast l1-Minimization Algorithms for Robust Face Recognition[J]. *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.
- [6]MA Rui, WANG Jiaxin, Song Yixu. Multimanifold Learning Using Locally Linear Embedding Nonlinear Dimensionality Reduction[J]. J Tsinghua University: Sci&Tech, 2008, 48(4): 583-586.
- [7]Xueping Chang, Zhonglong Zheng, Xiaohui Duan, Chenmao Xie. Sparse Representation-based Face Recognition for One Training Image per Person[C]. Lecture Notes in Computer Science, 2010, Volume 6215: 407-414.
- [8]Guijun Dan. Single-sample Face Recognition Algorithm Based on Virtual Sample Expansion Method[J]. Science Technology and Engineering, 2013, 13(14): 3908-3911.
- [9] Gao Q, Zhang L, Zhang D. Face recognition using FLDA with single training image per person, Applied Mathematics and Computation, 2008; 205( 2) : 726—734.
- [10]J. Lu, Y. Tan and G. Vang, “Discriminative Multimanifold Analysis for Face Recognition from a Single Training Sample per Person,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no.1, pp. 39-51, Jan. 2013.
- [11]M. Nabipour, A. Aghagolzadeh and H. Motameni, “A DCT-based multimanifold face recognition method using single sample per person.” the 8th Symposium on Advances in Science and Technology (8thSASTech), Mashhad, Iran, Feb 2014.
- [12]Mehrasa Nabipour, Ali Aghagolzadeh, Homayun Motameni. Multimanifold analysis with adaptive neighborhood in DCT domain for face recognition using single sample per person. The 22nd Iranian Conference on Electrical Engineering (ICEE 2014), May 20-22, 2014, Shahid Beheshti University.