

Robust Face Recognition Via Gabor Feature and Sparse Representation

Yu-Juan HAO^{1,a}, Li-Quan ZHANG¹

¹ College of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China

^a Corresponding author Yu-Juan HAO. E-mail address: 1243476593@qq.com

Abstract. Sparse representation based on compressed sensing theory has been widely used in the field of face recognition, and has achieved good recognition results. but the face feature extraction based on sparse representation is too simple, and the sparse coefficient is not sparse. In this paper, we improve the classification algorithm based on the fusion of sparse representation and Gabor feature, and then improved algorithm for Gabor feature which overcomes the problem of large dimension of the vector dimension, reduces the computation and storage cost, and enhances the robustness of the algorithm to the changes of the environment. The classification efficiency of sparse representation is determined by the collaborative representation, we simplify the sparse constraint based on L1 norm to the least square constraint, which makes the sparse coefficients both positive and reduce the complexity of the algorithm. Experimental results show that the proposed method is robust to illumination, facial expression and pose variations of face recognition, and the recognition rate of the algorithm is improved.

1 Introduction

Face recognition has the advantage of convenience and safety, easy to accept, not easy to counterfeit, which are widely used in the identification, video retrieval, safety monitoring and other fields. It is a hot research topic in the field of pattern recognition and artificial intelligence. In recent years, with development of compressed sensing theory and the L1 norm optimization technology, sparse representation has received the attention of many scholars at home and abroad.

In sparse representation, a signal can be represented as the most sparse linear combination of a given dictionary atom. Therefore, sparse representation has been widely used in image denoising and recovery, sparse representation has been widely used in pattern recognition and face recognition. In 2009, Wright John applied sparse representation to the field of face recognition, and achieved good recognition results in the face image under the condition of noise pollution or partial occlusion [1]. The principle of sparse representation is to use the dictionary to process the signal processing [2]. The signal represents a linear combination of a small number of atoms. Many scholars have done researches and improvement based the sparse representation algorithm.

Yang discuss the rationality of L1 norm solution of SRC and the validity of the classification method [3]. [6] believe that collaborative representation is more important than the L1 norm in the effective face recognition, and proposed a collaborative representation classification method (CRC), which is more efficient in computing. However, CRC does not provide a method to remove the image noise, so there is still a lack of recognition. Based on kernel principal component analysis (KPCA) and Fisher linear discriminant analysis (LDA), a complete kernel Fisher discriminant analysis (CKFD) algorithm is proposed by Yang [4]. Cevikalp propose kernel discriminant common vector method. The local tangent space alignment algorithm (LTSA) is proposed by [9], Zhang [10] propose an adaptive manifold learning based on LTSA. Lai [11] put up a sparse 2D local discriminant projection (S2DLDP) algorithm. [7][8] raise discriminant analysis of robust feature extraction of principal component analysis based on non parametric maximum entropy principle. Yang proposed GSRC (based SRC Gabor-feature) method, which uses Gabor features significantly reduce the size of the block dictionary and improve the rate of face recognition [12].

Huang [16] propose a neighborhood preserving embedding algorithm, the proposed algorithms not only

solve the weakness of traditional linear methods such as PCA, which are difficult to maintain the nonlinear flow of the original data, but also solve the shortcomings of the nonlinear method which are hard to obtain the low dimensional projection of the new sample points. Chen [17] present a local discriminant embedding (LDE) method for popular learning and pattern classification. Yan [18] propose a graph embedding framework to develop a new dimension reduction algorithm. Cai [19] propose a subspace method of spatial smoothing for face recognition and orthogonal Laplacian faces face recognition [21]. Fan [20] raise an improved LDA framework (LLDA) that can effectively capture the local structure of the samples.

Based on the current level of technology development, L1 norm sparsity determines SRC method whether it can be successfully applied to the problem of face recognition. And for this issue, [13] analysis and experiment the basic principle of SRC, and propose that SRC is more important than L1 norm sparsity in the classification methods based SRC, Zhang proposed a classification method based on sparse representation and rule of least squares SRC-RLS [14] [15]. Compared with the CRC method [6], (SRC-RLS) the results of the recognition is not only much better, but also can reduce the complexity of calculating. However, the SRC-RLS method directly uses the gray characteristics of the image recognition, so when the situation changes of light, posture and facial expressions, the accuracy of this recognition method will not be very high.

In the second part, we introduced Gabor feature extraction principle and our improved method. We lists the data of the method, in the third part, we introduce the sparse representation theory and its shortcomings, and point out that the sparse coefficients contain negative values which affect the result of the experiment, and introduces the collaborative representation. Finally, we validate the effectiveness of our method on a database which can improve the robustness of face recognition.

This paper proposes a new face recognition algorithm based on a improved Gabor feature and sparse representation (GSRC), and the least square method is used to replace the L1 solution. After the feature extraction of face recognition image, we fully constitute a complete dictionary of facial features, and then fuse the sparse representation algorithm, which will improve the recognition rate.

2 Gabor Feature Extraction

2.1 Gabor Filter Principle

First, Gabor function is obtained through the Fourier transform, the spectrum feature of Fourier transform is only in the infinite time domain which is difficult to fit for the local signal frequency analysis. A part of the time

domain signal is multiplied with the Gauss window function and the Fourier transform is performed, we can get Gabor transformation. It was first proposed by Gabor Dennis, in 1946. The Gabor transform will be defined as:

$$G_f(\omega, b) = \int_{-\infty}^{\infty} e^{-i\omega t} g_a(t-b) f(t) dt \quad (1)$$

The function of Gabor filter in the image processing is to achieve the input signal wavelet transform [17], so as to obtain the image of the local texture characteristics:

$$I_{mlpq} = \iint f(x, y) \varphi_{ml}(x - p\Delta x, y - q\Delta y) dx dy \quad (2)$$

$$(m = 0, \dots, M-1, l = 0, \dots, L-1)$$

M represents directions, L represents the scales; p represents the x -coordinate of the image pixels, q represents the ordinate of the image pixels; Δx and Δy are spatial sampling interval, usually set $\Delta x = \Delta y = 1$, $\varphi_{ml}(x, y)$ is the parent wave:

$$\varphi_{ml}(x, y) = a^{-m} \varphi(\tilde{x}, \tilde{y}) \quad (3)$$

In Eq.(3),

$$\tilde{x} = a^{-m}(x \cos \theta + y \sin \theta), \tilde{y} = a^{-m}(-x \sin \theta + y \cos \theta)$$

Direction change is defined as

$$\theta = l\Delta\theta$$

$$\Delta\theta = 2\pi / L \quad (4)$$

θ can choose different angles with the same spacing in 2π , thus making the filter in different directions have the same size of energy.

$$g(x, y) = \left(\frac{1}{2\pi\delta_x\delta_y} \right) \exp \left[-\frac{1}{2} \left(\frac{x^2}{\delta_x^2} + \frac{y^2}{\delta_y^2} \right) + 2\pi j W_x \right] \quad (5)$$

W represents frequency bandwidth. When $W = 0.5$, the filter can match the visual system. The frequency characteristic of Gabor wavelet can be realized in the space of u axis and v axis:

$$\psi(u, v) = \exp[-2\pi^2(\delta_x^2 u^2 + \delta_y^2 v^2)] * \delta(u - W)$$

$$= \exp \left(-2\pi^2 \left(\delta_x^2 (u - W)^2 + \delta_y^2 v^2 \right) \right) \quad (6)$$

$$= \exp \left(-\frac{1}{2} \left(\frac{(u - W)^2}{\delta_u^2} + \frac{v^2}{\delta_v^2} \right) \right)$$

$$\text{Frequency variance } \delta_u = \frac{1}{2\pi\delta_x}, \delta_v = \frac{1}{2\pi\delta_y}$$

However, the output of Gabor filter in high frequency state is relatively small, and it is necessary to change the basis function calculation method.

For people's facial features, there is also a unique characteristics of unity. In order to separate people, it should make full use of the unique characteristics of these key features, including the texture distribution of facial organs, geometric distance and the crowd, which is also the main theoretical basis for face recognition. We use the "three vertical and five horizontal" law, face detection in the vertical direction is divided into three parts. In the horizontal direction, the face is divided into five equal parts, each is equal to the width of a eyes.

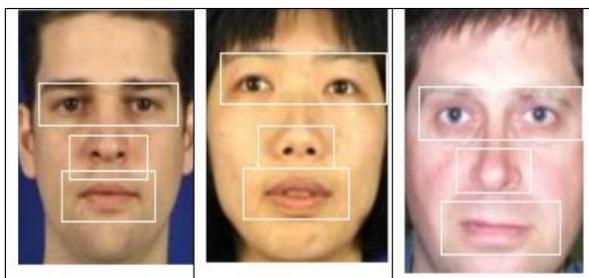


Figure1. Facial Organ Location. Figure 1 Shows the Location of the Eyes, Nose and Mouth.

We chose 40 filters to filter the image which contains 8 directions and 5 scales. The characteristics described by each filter is different. In order to be more intuitive to identify these differences, we intercepted a few filtered images were analyzed.



Figure2. Filtered Image It Shows The Characteristics Of Different Features After Filtering

In the face images, we need extract the key feature points, we choose the distance between each feature points as the feature vector of the face. Such as Fig.3.

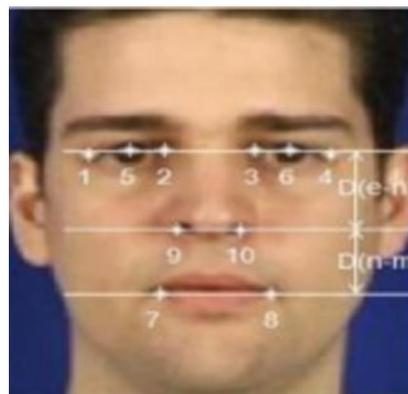


Figure3. Feature Map of Geometric Distance

In figure 3, we show the process of the feature extraction. Eye feature points: in the Cb image, the eye section shows a high pixel value, in Cr image, the eye section shows a low pixel value. Considering the character and select the appropriate threshold which can separate the eyes into a black area (the eye) and two white areas. Finally, we can get 6 feature points of the eyes: it is easy to get the feature points of the mouth, because the size of the pixel values in the Cr component has a great relationship with red. The red lips make it more in the pixel in the Cr component, then the sobel filter can be used to extract its contour. According to this inherent attribute, we can roughly determined in the image of the nose area. For the extraction of the feature points of the nose, it also can be expressed by the Sobel filter, and then extract the points on both sides of the nose root as a feature point.

In the research of the face recognition system, we need use the key features of the human face organs, and the feature points are used to represent the human face organs.

2.2 Improved Gabor feature extraction method

Based on the Gabor function, principle and design theory of filter, we have made some changes in the selection of Gabor kernel function, so as to be conducive to the implementation of the program. The two-dimensional Gabor kernel function we choose are shown in [18][19].

$$\begin{cases} W(x, y, \theta, \lambda, \phi, \sigma) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \phi\right) \\ x' = x\cos(\theta) + y\sin(\theta) \\ y' = -x\sin(\theta) + y\cos(\theta) \end{cases} \quad (7)$$

(x, y) represent a pixel coordinates in the diagram
 $\theta, \lambda, \phi, \sigma$ The parameters of $\theta, \lambda, \phi, \sigma$ are shown in Table 1.

Table 1.Data of $\theta, \lambda, \phi, \sigma$

parameters	Symbol	Set value
direction	θ	$\left\{ 0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8}, \frac{7\pi}{8} \right\}$
wavelength	λ	$\{4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\}$
phase	ϕ	$\left\{ 0, \frac{\pi}{2} \right\}$
Gauss radius	σ	$\sigma = \lambda$

Direction θ : Represent the direction of parallel stripes in the Gabor function. Wavelength λ : In image processing, it is said that the unit is a pixel, there is a certain range of needs less than the size of the image 1/5, the minimum is 2. Phase ϕ : In the range of -180 degrees to 180 degrees, Gauss radius σ : Represent the standard deviation of Gauss factor in the Gauss's function used in Gabor transform, $\sigma = \lambda$.

A set of filters consists of 5 different spatial frequencies and 8 different directions, which make up 40 different filters, as shown in Figure 4.

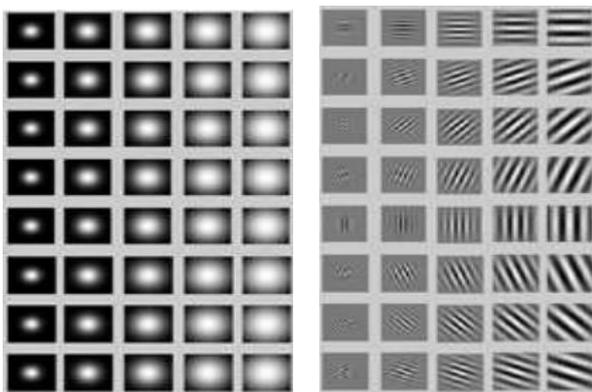


Figure 4 .Real part of filter (left) and imaginary part(right).

When we use this set of filters to deal with a face image, the number of processed images is equal to the number of filters. If these filters are applied to each pixel in the image, the filtered vector dimension is very large (proportional to the size of the image). So, this will lead to a huge amount of computing and storage costs, which will consume a lot of time to classify and identify. Our method is only applied to the filter to extract the ten benchmark points, to obtain the filter response. If the number of filters

is M , then the vector dimension of the filter response is $10 * M$. The number of filters corresponding to different values of θ is shown in Table 2.

Table 2 Numbers of Filter and the Values of θ

Numbers of filter	Values of different directions (θ)
5	$\{0\}$
10	$\left\{ 0, \frac{4\pi}{8} \right\}$
20	$\left\{ 0, \frac{2\pi}{8}, \frac{4\pi}{8}, \frac{6\pi}{8} \right\}$
30	$\left\{ 0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8} \right\}$
40	$\left\{ 0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8}, \frac{7\pi}{8} \right\}$

3 Sparse Representation Robust Recognition

3.1. Sparse Representation Theory

The essence of sparse representation is the decomposition of signals under the constraint of sparse regularization. The basic model that natural signal can be expressed as a linear combination of atomic signal predefined, and the combination coefficients is sparse, i.e. most of the coefficients is zero or close to zero.

In the face recognition technology based on sparse representation, each image size is $w \times h$, which can be regarded as a point of m dimensional space, $m = w \times h$, The information of different brightness, expression and occlusion of the same face are expressed in the low dimensional linear subspace. We define n_i pictures, n_i training samples of i subject form a matrix, $A_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n}] \in R^{m \times n_i}$. Any of the test samples ($y \in R^m$) of the class are similar to the linear subspace in which the columns are formed by the A_i .

$$y = a_{i,1}v_{i,1} + a_{i,2}v_{i,2} + \dots + a_{i,n_i}v_{i,n_i} \quad (8)$$

$a_{i,j} \in R$ is combination coefficient, $j = 1, 2, \dots, n_i$.

Because the class of the test sample can not be predicted in advance, a matrix A is defined for the whole training sample set, it is regarded as a series of k classes of training samples. All n_i are added together: $n = n_1 + n_2 + \dots + n_k$,

we regard n columns as vector A :
 $A = [A_1, A_2, \dots, A_k] = [v_{1,1}, v_{1,2}, \dots, v_{k,n_k}]$,

y can be expressed as a linear combination of the training sample set:

$$y = Ax_0 \in R^m \quad (9)$$

$x_0 = [0, \dots, a_{i,1}, a_{i,2}, \dots, a_{i,n_i}, 0, \dots, 0]^T \in R^n$, x_0 is coefficient vector, Non - zero items of x_0 should be corresponding to the class i of the training set. At this time the problem is transformed into solving equations $y = Ax$. Under normal circumstances, the equation set is not under determined, and its solution is not unique, but it can be solved by the following optimal problem: (l^0):

$$\hat{x}_0 = \arg \min \|x\|_0 \text{ subject to } Ax = y \quad (10)$$

Symbol $\|x\|_0$ express the number of nonzero elements in vector x . We can prove that for an arbitrary set of equations, when the number of nonzero elements of one of the solutions x is less than $m/2$, x is the most sparse solution $\hat{x}_0 = x$. However, the problem of finding the sparsest solution of an under determined system of linear equations is NP-hard, we need alternative ways to solve. By sparse representation and compression sensing theory, if the solution of X_0 is sparse enough, then the problem of l^0 norm can be replaced by the minimum l^1 norm: (l^1) $\hat{x}_1 = x \arg \min \|x\|_1$ subject to $Ax = y$ (11)

Due to the impact of noise or inadequate training samples and other reasons, the linear constraint in the optimization problem is not always established, the actual model can be modified as: $y = Ax_0 + z$, $z \in R^m$ represent noise, $\|z\|_2 < \varepsilon$. At the same time, sparse solution X_0 can still be achieved by solving the problem of l^1 norm:

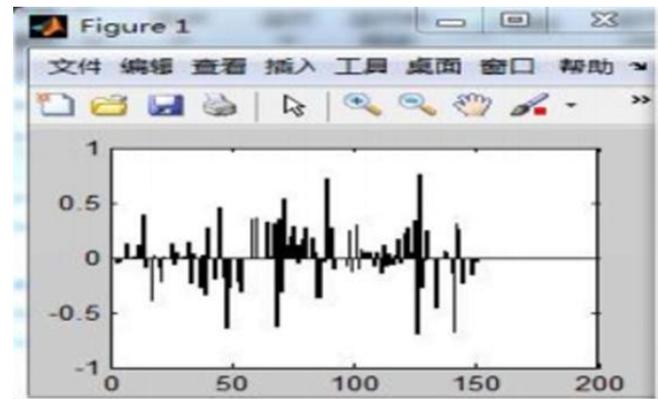
$$\hat{x}_1 = x \arg \min \|x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \varepsilon \quad (12)$$

We choose the ORL face database to illustrate the shortcomings of sparse representation. We select an image of subject 8, which have glasses occlusion from the test images to detect, whose mouth posture changes, the recognition result is not correct caused by linear conversion errors, the results as shown in figure 5 (a). From the figure 5 (b) and (c) show that the maximum coefficient does not belong to the correct subject by the classical SRC method. The error of training images detected although sparse coefficient is the largest, but it also has a negative

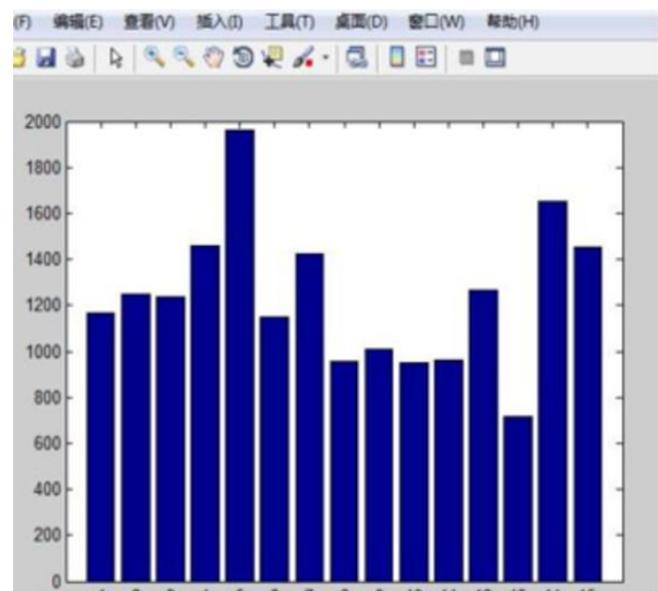
correlation coefficient absolute value maximum, other values are mostly negative. So the effect is not obvious.



(a)



(b)



(c)

Figure 5. Classical Sparse Representation. We Choose the Orl face Database Which Contains 40 People, Each Person Has

11 images. We select 15 people, every has 10 train images, one as the test image. The images are dealt with 12*8 by down sample. (a) the recognition result, test image belongs to subject 8, the result shows it belongs to subject 13. (b) the sparse coefficients. (c) the residuals of 15 people.

Feature obtained by the down sampling is not obvious, it also can not express the face feature. And the sparse coefficients of the traditional sparse said method are negative, which has great interference in the category judgment, if the coefficient is positive, this problem can be optimized.

3.2 Face Recognition Algorithm Based on Collaborative Representation

Many documents have highlighted the importance of sparse for classification, and have not studied the role of various training samples in the collaborative representation of test samples. Improve the accuracy of face recognition, a collaborative representation based classification (CRC) method is proposed in the literature [6]. In a practical face recognition system, in order to ensure the high recognition rate, feature dimension usually cannot be set too low. Therefore, there is no need for us to l^1 -norm. Taking into account the dictionary X may be under determined, we use l^2 -norm regularization. In order to use x to express the test sample y we propose the following collaborative representation of the regularized least squares method:

$$\hat{a} = \arg \min_a \left\{ \|y - xa\|_2^2 + \lambda \|a\|_2^2 \right\} \quad (13)$$

λ is regularization parameter, The function of l^2 -norm regularization $\|a\|_2$ is twofold. First of all, it makes the least squares solution stable, especially when x is less regular; secondly, it introduces a certain number of sparse to \hat{a} . However, this degree of sparsity is weaker than the l^1 -norm. We can get this:

$$\hat{a} = (x^T x + \lambda I)^{-1} x^T y \quad (14)$$

Order $P = (x^T x + \lambda I)^{-1} x^T$ Obviously, P is independent relative to y , We can calculate P as a projection matrix. Once the test sample y arrival, we will simply pass the P to the Py , which makes the calculation speed of the collaborative representation is very fast. Depending on the classification of \hat{a} in the CRC and the classification of the \hat{a} depending on the SRC are similar. In addition to using the

specified class of $\|y - X_j \hat{a}_j\|_2$ to classify the reconstructed residuals Where \hat{a}_j is corresponding to the class of coding coefficient vector, l^2 -norm sparse term can also bring some discriminant information. Therefore, we propose to use both of them in classification, so that the classification accuracy can be improved by the use of $\|y - X_j \hat{a}_j\|_2$.

3.3 Combination of Improved Gabor Features and Sparse Representation

When the images exist illumination, expression and pose variation, the recognition rate of traditional sparse representation will be decreased. In order to further improve the algorithm accuracy and robustness, We will combine feature extraction method which characterize the facial feature vectors including the Gabor filter in response to the $10 \times M$ value and 8 geometric distance parameters of the feature extraction method with sparse representation algorithm, and the least square method is used to solve the norm, and the condition of the sparse coefficient is positive.

According to the proposed extension of the Gabor feature definition, extracting all the Gabor features of the training sample set image, the corresponding Gabor feature set is extracted from the eyes, nose and lips, and set up the corresponding feature set. $X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,n_i}]$. The Gabor feature set for all classes of training samples is $X = [X_1, X_2, \dots, X_K]$, All the training samples are set to represent the test samples. That is to simplify the SRC algorithm based on norm of sparse constraints to the rule of least squares method,

$$(\hat{p}) = \arg \min_p \left\{ \|y - Xp\|_2^2 + \lambda \|p\|_2^2 \right\} \quad (15)$$

P is coefficient vector, λ is expressed rule parameter, Regularization term $\lambda \|p\|_2^2$ it makes the least squares solution stable, especially when X is less regular; secondly, it introduces a certain number of sparse to \hat{p} . P is independent relative to y , We can calculate P as a projection matrix.

Improved algorithm flow:

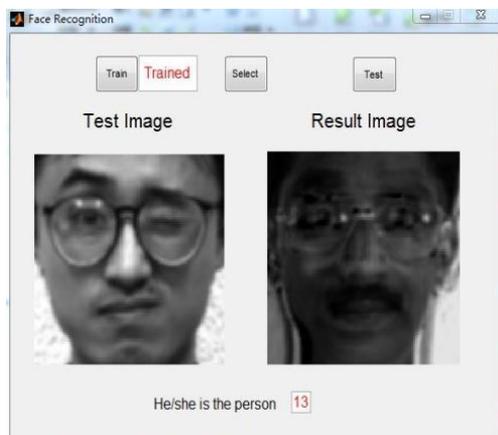
1. Extract the Gabor features of the three parts of the training sample set and the test sample, then respectively get X and Y ;
2. The test sample feature vector y is based on the training sample set X collaborative representation coding $\hat{p} = Py$, The projection matrix is

$$P = (x^T x + \lambda I)^{-1} x^T;$$

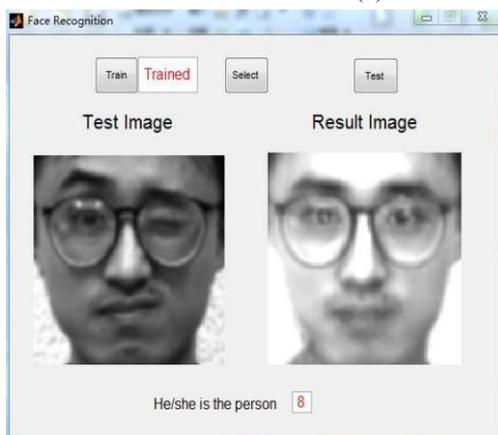
3. calculate regular residual $r_i = \|y - X_i \hat{p}_i\|_2 / \|\hat{p}_i\|_2$, \hat{p}_i is the encoding coefficient vectors for class i ;
4. Output: $identity(y) = \arg \min_i \{r_i\}$

4 Simulation and Results

In order to verify the effectiveness of the algorithm, we conduct a simulation experiment, we select two standard face recognition databases, that is, the performance comparison test of the proposed algorithm is carried out on Extended Yale B database and AR database. In this paper, we use the support vector machine classification algorithm (SVM) and the classical SRC algorithm to compare with the improved algorithm. Compare with the Fig.5, we see the improved result.



(a)



(b)

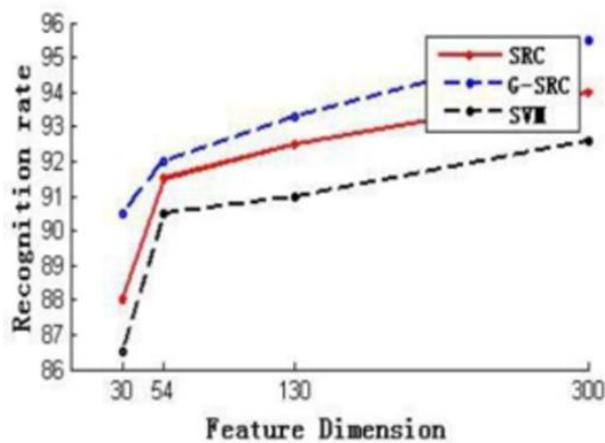
Figure6. Compare the Improved Result and the Src, the Progress is the Same as fig.5.(a) Represent the Result of Fig.5,(b)Shows the Correct Result by Our Method.

4.1 Face Experiment on Extended Yale B

The Extended Yale B face database contains 38 individual frontal face image, each in different illumination conditions have 64 pictures, the size of each photo is 192 168. We randomly selected 32 samples as training samples, and the rest as the test samples. Like the classical sparse representation, we select the feature space dimension is 30,54,130,300, Fig.8 Recognition rate of three kinds of classifiers in four kinds of feature dimension.



(a)



(b)

Figure 7. Recognition Rate on Extended Yale b Database, (a)Test Sample Images.(b)Recognition Rate of Three Kinds of Classification Methods.

We can see that the improved method has greatly improved the recognition rate compared with the original method. The feature space dimension is 300, our method can achieve 95.3%, the SRC is 94.1%.

4.2 Face Experiment on AR Database

AR face database contains a total of 4000 images collected from 126 volunteers, including the changes in light, angle, facial expression and occlusion. We select 50 classes, each class of 26 samples, the image size is 165 120. We choose 14 images per class as a training sample, and the others are used as test samples. The

feature space dimension is 30, 54, 130, 300. Fig.9 The experimental results obtained by the three algorithms with the change of dimension.

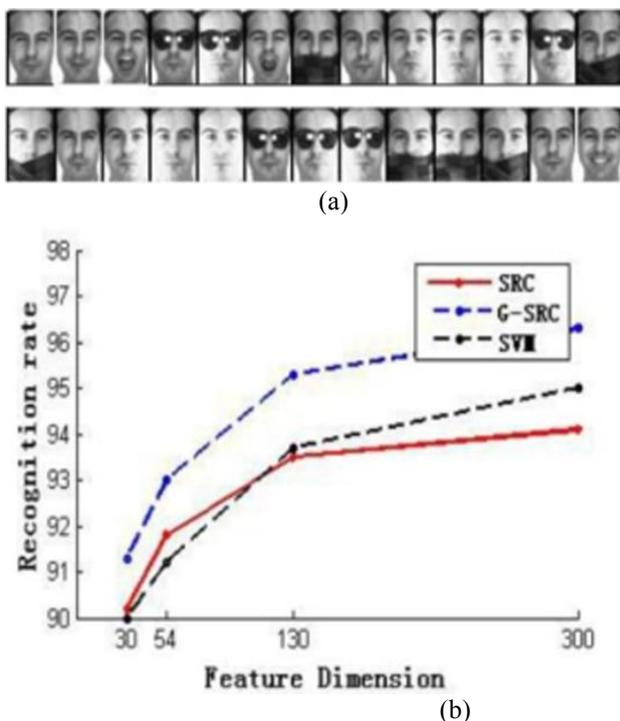


Figure 8. Recognition Rate on Ar Database. (a) Sample of Someone. (b) the Recognition Rate of Three Kinds of Classification Methods.

We can see that the improved method has greatly improved the recognition rate compared with the original method. The feature space dimension is 300, our method can achieve 96.3%, the SRC is 94.1%.

Acknowledgments

In the research of image recognition, feature extraction can be described as an essential step. The advantage of feature extraction determines the accuracy of the recognition. We propose an improved Gabor feature extraction algorithm to obtain the accurate position of the eyes and mouth and extract features; Using Sobel edge detection technology determines the position of the nose and extract the feature points.

We calibrate the facial feature points, then apply Gabor filter at each characteristic point respectively. The purpose of this process is to extract the local details of the facial organs.

The Gabor feature extracted by this method is not only low in dimension, but also has good robustness. Then the least square method is used to solve the solution, which ensures the non-negativity of the coefficients. The experimental results show that the new method is robust to illumination, occlusion and other changes.

References

1. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31 210-227(2009)
2. J. Z. Huang, X. L. Huang, D. Metaxas Simultaneous image transformation and sparse representation recovery. In: *Proceedings of the 26th IEEE Conference on Computer Vision and Image Recognition*. Anchorage United States. 1-8(2008)
3. J. Yang, L. Zhang, Y. Xu, and J. Y. Yang, Beyond sparsity: The role of L1-optimizer in pattern classification, *Pattern Recognit*, 45 1104-1118(2012)
4. J. Yang, A. F. Frang, J.-Y. Yang, D. Zhang, and Z. Jin, KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 27(2) 230-244(2005)
5. H. Cevikalp, M. Neamtu, and M. Wilkes, Discriminative common vector method with kernels, *IEEE Trans. Neural Netw.* 17(6) 1550-1565(2006)
6. L. Zhang, M. Yang and X. C. Feng, Sparse representation or collaborative representation: Which helps face recognition? In *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 3 7(2011)
7. R. He, W.-S. Zheng and B.G. Hu, Maximum Correntropy criterion for robust face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 1561-1576(2011)
8. R. He, B.-G. Hu, and X. T. Yuan, Robust discriminant analysis based on non-parametric maximum entropy, In *Proc. Asian Conf. Mach. Learn*, 120-134(2009)
9. Z. Zhang, H. Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Scientific Computing*, 26 313-338(2004)
10. Z. Y. Zhang, J. Wang, H. Y. Zha. Adaptive manifold learning *IEEE. Trans. Pattern Anal. Mach. Int.* 32 253-265(2012)
11. Z. Lai, M. Wan, Z. Jin, and J. Yang, Sparse two-dimensional local discriminant projections for feature extraction, *Neuro.Computing*, 4 629-637,(2011)
12. M. Yang, L. Zhang, Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary, In *ECCV*, 448-461 (2010)
13. N. Zhang, J. K. Yang. Nearest neighbor based local sparse representation classifier, In: *Proceedings of the 2010 Chinese Conference on Pattern Recognition*. Chongqing, China: CCPR, 400-404 (2010)
14. R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B*, 73(3) 273-282(2011)
15. J. Zhang, R. Jin, Y. M. Yang, Y. M. Hauptmann, A. G. Modified, Logistic regression: an approximation to SVM and its applications in large-scale text

- categorization, In: Proceedings of the 20th International Conference on Machine Learning. Washington, United states: ICML, 888-895(2003)
16. Y.FU, T.S. Huang. Graph embedded analysis for head pose estimation, IEEE International Conf. on Automatic Face and Gesture Recognition. (2006)
 17. H .T. Chen, H. W. Chang, and T. L. Liu, Local discriminant embedding and its variants, Proc. Conf. Comput.Vis. Pattern Recognit. 846–853(2005)
 18. S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, Graph embedding and extension: A general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29, 40–51(2007)
 19. D.Cai, X. He, Y. Hu, J. Han and T. Huang, Learning a spatially smooth subspace for face recognition, Proc. Comput.Soc. Conf. Comput.Vis. Pattern Recognit.,1–7 (2007)
 20. Z. Fan, Y. X u, and D. Zhang, Local linear discriminant analysis framework using sample neighbors, IEEE Trans. Neural Netw,22,1119–1132(2011)