Competitive Representation Based Classification Using Facial Noise Detection

Tao Liu
Chongqing Key Laboratory of Computational Intelligence
College of Computer Science and Technology, Chongqing University of Posts and Telecommunications
Chongqing, China

Ying Liu
Chongqing Key Laboratory of Computational Intelligence
College of Computer Science and Technology, Chongqing University of Posts and Telecommunications
Chongqing, China

Cong Li
Chongqing Key Laboratory of Computational Intelligence
College of Computer Science and Technology, Chongqing University of Posts and Telecommunications
Chongqing, China

Chao Li
Chongqing Key Laboratory of Computational Intelligence
College of Computer Science and Technology, Chongqing University of Posts and Telecommunications
Chongqing, China

***Abstract*—Linear representation based face recognition is hotly studied in recent years. Competitive representation classification is a linear representation based method which uses the most competitive training samples to sparsely represent a probe. However, possible noises on a test face image can bias the representation results. In this paper we propose a facial noise detection method to remove noises in the test image during the competitive representation. We compare the proposed method with others on AR, Extended Yale B, ORL, FERET, and LFW databases and the experimental results show the good performance of our method.**

Keywords—face recognition; sparse representation; biometrics; noise detection

I. Introduction

Face recognition (FR) technology has been attracting many researchers to study so far [1,2].In recent years, a lot of FR methods using linear representation are proposed [3,4].There are at least two obvious merits of linear representation based methods: First, a test sample synchronously matches with all training samples from gallery set, which guarantees high recognition accuracy. Second, it is verified that the recognition results are not sensitive to the type of feature, which means very simple feature such as down-sampled images can serve well. In a linear representation based FR method, an assumption is made that samples from the same class are distributed in a same subspace, which means a given test image $\mathbf{y}$ can be linear represented by a training set $\mathbf{A}$ as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{1}$$

where $\mathbf{x}$ is a coding vector and in this model images are transformed from matrix into a vector. In classical biometric methods a test sample need to match each training sample one at a time, but we can see from (1) that the correlation between the test sample and all training samples is built in one linear representation model. It is quite common that the training set matrix $\mathbf{A}$ is over-complete, so that a sparse representation scheme is proposed to enforce only a few training samples to respond in the representation model. One way to implement the sparse representation scheme is to introduce a l_0 -norm constraint on a coding vector, which is formulized as:

$$\min \|\mathbf{x}\|_0 \quad s.t. \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le \varepsilon \tag{2}$$

where $\|.\|_0$ denotes the l_0-norm, which counts the number of nonzero entries of the coding vector and ε is a small error tolerance. Since to solve the model (3) is a NP-hard problem, in [5] a new model was built as

$$\min \|\mathbf{x}\|_1 \quad s.t. \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le \varepsilon \tag{3}$$

where $\|.\|_1$ is the l_1 -norm of a vector, which sums up all absolute values of entries in the vector. It is proved that under some conditions the solution of (3) is a good approximation of (2). To optimize (3), a further equal transformation is made to link the sparse representation model to the Lasso regression, which is

$$\min \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda \|\mathbf{x}\|_1 \tag{4}$$

where λ is used to balance the l_1 -norm based regularization term and the l_2 -norm based fidelity term.

Many follow-up studies improved the performance of the recognition using (3) in their sparse representation FR methods. In [6], robust sparse coding (RSC) methods is proposed which is robust to various kinds of noises. In RSC, the fidelity term is replaced with sigmoid-like function, which makes the representation less sensitive to outlier pixels. In [7], a sparse representation model is proposed based on the maximum correntropy criterion, which is also a method using a more robust function to replace the l_2 -norm based fidelity term in (4). But, furthermore, a new technique called half-quadratic framework is proposed in this work, which improves the sparse representation in terms of both error correction and error detection [8].

Some researchers argue that without imposing sparsity on coding vector can still obtain good classification accuracy. In

[9], linear regression models are used to represent a probe one class at a time. And their later work uses Huber estimator to achieve more robust regression against different levels illumination changes. In [10],authors suggest that to use ridge regression instead of Lasso regression model is good for the case of Gaussian distributed noise and introduce the collaborative representation based classification (CRC) with the non-sparse l_2 -norm to regularize the representation coefficients. However according to the investigation in [11] to restrict the number of training samples in the collaborative representation can increase the performance of CRC further. Inspired by this study, a two-phase collaborative representation method is proposed, in which the first phase is to choose a subset of training samples which is close to the test sample and the second phase is to conduct CRC on the chosen training samples. The two-phase collaborative representation can be considered as another way to realize sparse representation in the sense that only a few training samples are involved in final representation. This kind of sparse representation scheme is called supervised sparse representation because the sparsity comes from the supervision in representation itself not l_1 -norm based regularization. In [12], experiments are conducted to show that to use several rounds of representation to subtly choose the training set is better than one-shoot training sample picking. It is not surprising that multiple rounds of training sample picking is quite time-consuming and however a fast calculation method, called competitive sparse representation classification (CSRC), is proposed in [13]. The method only deletes the lowest competitive samples in each round of representation and uses a fast algorithm to avoid repeatedly calculating the matrix inverse in each round.

However the supervised sparse representation has not considered the possible massive noises on a probe such as scarf occlusion so far. The noise could harm the procedure of picking the competitive training sample subset so that the chosen samples show the appearance of bias in favor of the occlusion. In this paper, we propose a new competitive sparse representation classification using facial noise detection (CSRC-FND). We do not only employ the multiple round of representation to select the most competitive subset of training samples, but also to delete the possible outlier pixels.

The paper is organized as follows: In section 2, we first introduce the competitive sparse representation classification method and then we propose our CSRC-FND method. In section 3, an analysis of the proposed method is given by illustrating the difference between CSRC and CSRC-FND. We conduct several experiments in section 4. And the conclusion is made in Section 5.

II. The Proposed Method

A. Competitive representation classification

We first briefly introduce the competitive representation classification method. $\mathbf{A}\in R^{m\times n}$ denotes a subset of training samples, where each column is an training sample with m pixels and n is the number of training samples in the subset. The procedure of competitive representation is multiple rounds of representation of a probe sample $\mathbf{y}\in R^m$ and in each round of representation the least competitive training samples are removed from the subset $\mathbf{A}$. For the first round, n equals the total number of training samples. To implement the competitive representation, the following ridge regression model is used:

$$\mathbf{x}=\arg\min_{x}\left\{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\lambda\|\mathbf{x}\|_2^2\right\} \tag{5}$$

where λ balances the regularization term and the fidelity term. The coding vector $\mathbf{x}$ has the close form solution:

$$\mathbf{x}=(\mathbf{A}^T\mathbf{A}+\lambda\mathbf{I})^{-1}\mathbf{A}^T\mathbf{y} \tag{6}$$

where $\mathbf{I}$ is an identity matrix. In each round, the coding vector $\mathbf{x}$ is calculated according to the subset $\mathbf{A}$. Then we need to sort the absolute value of entries of $\mathbf{x}$ and remove the samples associating with the smallest the absolute value (the number may be more than 1) from the subset $\mathbf{A}$. We repeatedly do the competitive representation until the subset $\mathbf{A}$ is small enough.

In the final decision, (5) is used to find the coding vector again. Then the decision is made by

$$\begin{aligned}&\mathrm{ID}(i)=\arg\min_i d_i(\mathbf{y})=\arg\min_i \|\mathbf{y}-\hat{\mathbf{y}}_i\|_2\\&\hat{\mathbf{y}}_i=\sum_j a_i^j x_i^j\\&i=1,2,\ldots,c\end{aligned} \tag{7}$$

where c is the total number of classes, $\hat{\mathbf{y}}_i$ is the prediction of the test sample by the i th class, a_i^j is a training sample of the i th class in the final subset and x_i^j is the corresponding coefficient in the coding vector.

B. New method using facial noise detection

The competitive representation classification method performs very well in the cases where noises on test samples are not strong. However in some applications of FR faces can be contaminated by severe variations such as massive partial occlusion and pixel corruptions. These noises can make the procedure of competitive representation bias, which means the chosen training samples show preference for matching the noise so that the competitiveness of the genuine samples would be suppressed. Inspired by supervised sparse representation itself, we consider to combine the noise detection with the competitive representation procedure. Hence in each round of the competitive representation of a test sample we not only pick training samples with high competitiveness but also to identify some possible contaminated pixels. To facilitate the presentation of our method, a unified model is used to implement the idea given as:

$$\arg\min_{x}\|\mathrm{Diag}(\mathbf{w})(\mathbf{y}-\mathbf{A}\mathbf{x})\|_2^2+l\,\|\mathrm{Diag}(\mathbf{d})\mathbf{x}\|_2^2 \tag{8}$$

where $\mathrm{Diag}(g)$ is to transform a vector to a diagonal matrix. Using this model, we use index vector to identify active pixels and active samples. $\mathbf{w}\,\hat{\mathrm{I}}\,R^m$ is the index vector to specify which pixels are used in the model and $\mathbf{d}\,\hat{\mathrm{I}}\,R^n$ specifies which training samples participate the representation. The entries of $\mathbf{w}$ and $\mathbf{d}$ are binary:

$$w_i \text{ and } d_j = \begin{cases} 1 & \textit{active} \\ 0 & \textit{inactive} \end{cases} \tag{9}$$

To set the value of $\mathbf{w}$ and $\mathbf{d}$, it is easy to specify the active training samples and active pixels. The model (8) can be seen as a weighted representation model with binary weights and $\mathbf{w}$ is the weight for pixels and $\mathbf{d}$ is the weight for training samples. A good feature of model (8) is that it has close form solution as follows:

$$\mathbf{x} = (\mathbf{A}^T \mathrm{Diag}(\mathbf{w})\mathbf{A} + \lambda\, \mathrm{Diag}(\mathbf{d}))^{-1} \mathbf{A}\mathrm{Diag}(\mathbf{w})\mathbf{y} \tag{10}$$

At the beginning, we set the initial value of $\mathbf{w}$ and $\mathbf{d}$ to vectors with all entries equal to one. Then we start the competitive representation. In each round of representation we calculate the results by (10) and then we find the entries (associated with $d_i = 1$) in $\mathbf{x}$ which have the least absolute values and set corresponding values in $\mathbf{d}$ to 0 (the changed values will not ever change back again). At the same time we calculate the representation residual by $\mathbf{e}=\mathrm{Diag}(\mathbf{w})(\mathbf{y} - \mathbf{A}\mathbf{x})$, and find the entries in $\mathbf{e}$ (associated with $w_i = 1$) whose absolute values are greater than a preset small positive value ξ and set corresponding values in $\mathbf{w}$ to 0 (the values will not ever change back again). By doing so, we try to identify the contaminated pixels on the test sample. The reason why we can find out corrupted pixels is that clear pixels can be matched by the linear model well while contaminated pixels deviate the real grey scale, which causes a big absolute residual. In addition we limit the lowest percentage of number of pixels that are involved in the representation. That is to say the ratio of zero entries to all number of entries in $\mathbf{w}$ should be no more than r (which is preset).

The stop condition of competitive representation is that there are only the k most competitive training samples are active in the representation. That is to say, if there are only k entries in $\mathbf{d}$ are equal ones, we stop the representation phase and do final representation. At last, the decision is made as in CRC method:

$$\mathbf{x}^* = Diag(\mathbf{d})\mathbf{x} \tag{11}$$

where $\mathbf{x}^*$ is the final active coding vector.

$$\mathrm{ID}(i) = \arg\min_i \|\mathbf{y} - \mathbf{A}_i\mathbf{x}_i^*\|_2 \big/ \left(\|\mathbf{x}_i^*\|_2 + \tau\right) \quad i = 1,2,\ldots,c \tag{12}$$

where τ is used to prevent denominator equal to zero. So that the proposed method can be summarized as the following algorithm:

Algorithm: Competitive sparse representation classification using facial noise detection (CSRC-FND)

Input: a probe image $\mathbf{y} \in R^n$, the initial dictionary $\mathbf{A} \in R^{m\times n}$.

Initial value: $\mathbf{w} = [1,1,\ldots,1] \in R^m$ and $\mathbf{d} = [1,1,\ldots,1] \in R^n$

Repeat

Compute the coding vector $\mathbf{x}$ according to (10);

Set the entries of $\mathbf{d}$ associated with the least absolute values of $\mathbf{x}$ to 0;

if $\dfrac{sum(\mathbf{w})}{m} < r$

Set the entries of $\mathbf{w}$ associated with the entries of $\mathbf{e}$ that $e_j > \xi$ to 0,

$j = 1,2,\ldots,m$;

end;

Until $sum(\mathbf{d}) \le k$

Compute the coding vector $\mathbf{x}$ according to(10);

Compute the final active training samples according to (11) ;

Identify : $\mathrm{ID}(i) = \arg\min_i \|\mathbf{y} - \mathbf{A}_i\mathbf{x}_i^*\|_2 \big/ \left(\|\mathbf{x}_i^*\|_2 + \tau\right)$
$i = 1,2,\ldots,c$

Output: i (the identity of $\mathbf{y}$).

III. Experiments

In this section, the performance of CSRC-FND is validated by extensive experiments. All the experiments are carried out using MATLAB on 2014a on a desktop with 3.30GHz CPU and 16G RAM. The normalized images are used in all experiments.

A. *Parameter Setting*

The proposed method includes three parameters, i.e. the representation steps N which represents the iterations of algorithm, the number M of removed feature in every step, and the number S of removed sample in every step. The three parameters codetermine which features and samples can be retained to linearly represent the testing image. Generally, the fewer the number of the retained noise and the number of the retained interferential samples are, the higher recognition rate of our method is. Through many experiments and analysis, the proposed method can obtain a good performance when $N = 20$, $M = 0.03*m$, and $S = 0.04*n$.

B. *Face Recognition without Occlusion*

In this subsection, the performance of CSRC-FND is shown in FR without occlusion, such as posture variations and expression changes. Moreover, we compare the proposed method with LRC [9], CRC [10], SRC without expansion version [5], and CESR [7]. In addition, all experiments are performed on three public databases, namely, AR [14], ORL [15], FERET [16], and LFW [17].

1) *AR Database*

In the experiment, 2600 images of 50 male and 50 female are chosen from the AR database. Each subject contains 26 images that are divided into two sessions, i.e. session 1 and session 2. We use 7 images of session 1 and session 2 for training and testing (i.e. Fig. 1), respectively. Moreover, the resolution of the images is 50×40. The recognition results of CSRC-FND and the compared methods are shown in TABLE Ⅰ. CSRC-FND obtains the highest recognition rate among the several methods, i.e. 96.71%. CESR uses a robust loss function to reduce the effect of noise in representation, but its recognition rate is lower 5.42% than our method. In addition, the recognition rate of CRC which uses l_2-norm to loss function is also lower than CSRC-FND. That is, it is very effective that CSRC-FND removes some invalid pixels and samples for improving the recognition rate. Moreover, the recognition rate of CSRC-FND is 76.14%.

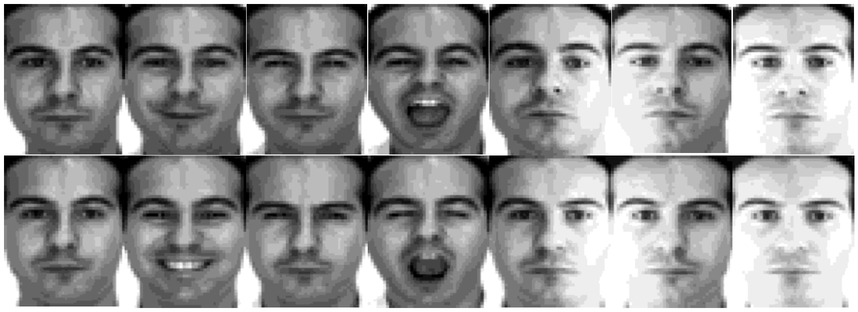


Fig. 1. The part of images of one subject in AR. Top: these facial images are the first seven images in session 1. Bottom: these images are the first seven images in session 2

TABLE I. THE RATES OF THE SEVERAL METHODS ON THE AR DATABASES

Methods	**LRC**	**CRC**	**SRC**	**CESR**	**CSRC-FND**
Recognition rates	76.14%	92.71%	75.71%	91.29%	96.71%

2) ORL Database:

ORL database contains 40 subjects and per subject consists 10 images with expression and posture changes, (i.e. Fig. 2). Each subject is separated into two parts: the first five images are used for training and the last five images are used for testing. All used images are cropped to 50×40. The in TABLE Ⅱ lists the recognition results of LRC, CRC, SRC, CESR and CSRC-FND. CESR and CSRC-FND obtain the highest recognition rate among all methods, i.e. 91.50% and 91.00%, respectively. In addition, the recognition rates of LRC (88%), CRC, and SRC are 3%, 5%, and 1.5% lower than our methods, respectively.

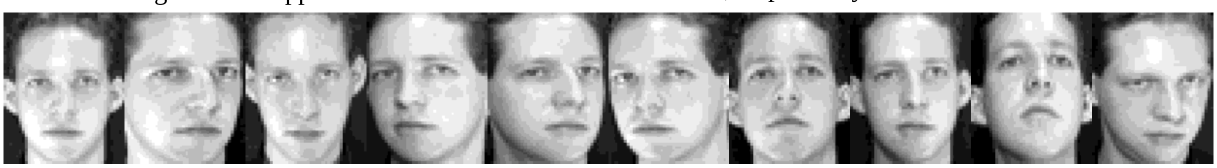


Fig. 2. The images of the first subject in ORL

TABLE II. THE RATES OF THE SEVERAL METHODS ON THE ORL DATABASES

Methods	**LRC**	**CRC**	**SRC**	**CESR**	**CSRC-FND**
Recognition rates	88.00%	86.00%	89.50%	91.50%	91.00%

3) FERET Database

FERET database consist of 1400 images of 200 subjects, i.e. each subject has 7 images, i.e. Fig. 3. These images mainly include posture variations. The first two images of per individual are using for training set, and the rest five images are used for testing. These images are resized to 40×40 in this experiment. From in TABLE Ⅲ, we can observe that the recognition rates of CESR and CSRC-FND are very close, the margin is only 0.1%, i.e. CESR is 65.90% and CSRC-FND 65.80%. LRC also obtains very high recognition rate, i.e. 64.10%. The performance of CRC is poor, the margin between it and our method is 17.20%. Therefore, CSRC-FND largely improves the performance of recognizing face image with posture variations by removing invalid features and samples. In addition, SRC obtains a good result, i.e. 61.50%.

Fig. 3. The seven images of the first subject in FERET

TABLE III. THE RATES OF THE SEVERAL METHODS ON THE FERET DATABASES

Methods	**LRC**	**CRC**	**SRC**	**CESR**	**CSRC-FND**
Recognition rates	64.10%	48.60%	61.50%	65.90%	65.80%

4) LFW Database

This database is a large-scale database of face photographs designed for unconstrained FR, which includes of pose variations, illumination variations, expression variations, misalignment variations and occlusion variations, and so on. Some examples are presented in Fig. 4. We gray the images in the LFW database and resize into 100×80 . We choose a subset which consist of 143 subjects with no less than 11 images each individual from LFW. Training set consists of the first ten images in each subject, and remaining images are used as testing. The recognition rates of the several methods are listed in TABLE Ⅳ. It is obvious that the margins between our method and the other several methods are very large the minima margin is 7.03%. Only is the recognition rate of CSRC-FND higher 60% among all methods. The recognition rates of LRC, CRC, SRC, and CESR are 40.09%, 55.14%, 24.31%, and 44.64%, respectively.

Fig. 4. A part of images in a specific subject from LFW

TABLE IV. THE RATES OF THE SEVERAL METHODS ON THE LFW DATABASES

Methods	**LRC**	**CRC**	**SRC**	**CESR**	**CSRC-FND**
Recognition rates	40.09%	55.14%	24.31%	44.64%	62.17%

C. Face Recognition with Occlusion

In this subsection, the robustness of methods is tested by dealing with the complex problems, such as corrupted facial, and occluded facial. And LRC, CRC, SRC, and CESR are compared with our method in these experiments. And the resolution of images in these experiments is 50×40.

1) FR with Pixel corruption

In this experiment, 1984 images of 31 individuals are chosen from the Extended Yale B database [18,19] and each individual has 64 images which are divided into 5 subsets according to different light intensities.

As in Fig. 5 (a), subset 1 includes the low light intensity and this subset is used for training, and subset 3 with high light intensity shown in Fig. 5 (b) is used for testing. Moreover, a certain percentage pixels in the testing are randomly chosen and replaced with random value in the range between 0 and the maximum value of pixel in the testing (Fig.5. (c)).

We list recognition results of LRC, CRC, SRC, CESR and CSRC-FND under different percentage of corruption in TABLE Ⅴ. We observe that performance of CSRC-FND is more stable than the other several methods under different corruption percentages. The recognition rate of our method occurs change until corruption percentage equals to 50%. However, the rates of the other methods are reduced when corruption percentage exceed 10%. It could illustrate that CSRC-FND can detect accurately the noise in testing facial image, and removing interfere samples is favor of face recognition.

(a) (b) (c)

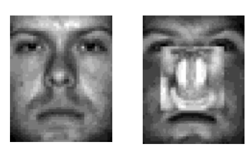
Fig. 5. Recognition under 40% random corruption. (a): An image in subset 1. (b): An image in subset 3. (c): A testing image with 40% random corruption

TABLE V. THE RATES (%) OF THE SEVERAL METHODS ON THE VERSUS DIFFERENT PERCENTAGE OF CORRPTUION

Corruption (%)	**LRC**	**CRC**	**SRC**	**CESR**	**CSRC-FND**
0	100.0	100.0	100.0	71.77	100.0
10	100.0	99.46	80.65	71.04	100.0
20	98.66	93.55	79.57	70.97	100.0
30	95.97	77.15	76.34	70.16	100.0
40	87.63	52.42	71.28	65.86	100.0
50	65.86	30.65	58.06	65.32	99.93
60	39.52	19.36	38.44	59.68	98.39
70	24.19	6.183	23.39	52.96	83.07

2) FR with Block Occlusion

In this experiment, we validate the performance that CSRC-FND recognizes the images with block occlusion. And LRC, CRC, SRC, and CESR are tested as the comparisons. As in previous experiment, the subset 1 in Extended Yale B is used for training and subset 3 for testing. A middle block in the testing image is replaced by an unrelated image, i.e. Fig.6 (b). The results of LRC, CRC, SRC, CESR and CSRC-FND under different the occlusion percentage from 10% to 60% are listed in TABLE Ⅵ. From the table, we can observe obviously that 10% occlusion is more serious than 20%, 30% occlusion for recognition. This is due to the fact that the value of the noise with 10% occlusion has larger different with the value of the fidelity pixel than 20%, 30% occlusion. However, the recognition rate of our method still equals to 100%. In addition, the performance of CSRC-FND is very stable, the margin between 10% and 50% is only 3.76%. Even the occlusion percentage reaches at 60%, its recognition rate is equal to 73.66%. It is more robust than CESR with using robust loss function. The performance of the other methods is also nonideal, i.e. their performance is very sensitive to the occlusion percentage.



(a) (b)

Fig. 6. Recognition under 30% block occlusion. (a) An image in subset 3. (b) A testing image with 30% block occlusion

TABLE VI. THE RATES (%) OF THE SEVERAL METHODS ON THE VERSUS DIFFERENT PERCENTAGE OF OCCLUSION

Occlusion (%)	**LRC**	**CRC**	**SRC**	**CESR**	**CSRC-FND**
10	68.28	37.10	54.83	68.55	100.0
20	75.27	57.80	59.95	63.98	100.0
30	77.69	47.31	63.44	64.25	100.0
40	70.67	43.01	60.21	55.91	99.46
50	68.55	43.01	54.84	52.15	96.24
60	47.85	30.11	40.05	37.90	73.66

IV. CONCLUSION

In this paper, a new competitive representation based FR method called CSRC-FND is proposed. In order to void the bias of choosing the training samples subset of high representation competiveness, the noise detection method is employed to remove the pixels which are very likely to have been contaminated. To implement both the competitive representation and noise detection, a weighted regression model is presented which involves two binary weight vectors that one is used to identify the active samples and another is to specify the active pixels in repeated competitive representations. According to the experiments, the proposed method outperforms the competitive sparse representation classification and show promising performance. However, the noise on facial image is very various. It is hard for CSRC-FND to remove the noise on the image and inference samples in training set, accurately. Therefore, we will focus on this problem in the future.

REFERENCES

[1] L. Luo, J. Yang, J. Qian, and Y. Tai, “Nuclear-L1 norm joint regression for face reconstruction and recognition with mixed noise,” Pattern Recognition, vol. 48, no. 12, pp. 3811-3824, 12, 2015.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” Acm Computing Surveys, vol. 35, no. 4, pp. 399-459, Dec, 2003.

[3] Y. Xu, Q. Zhu, Z. Fan, 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, no. 0, pp. 138-148, 2013.

[4] J.-X. Mi, J.-X. Liu, and J. Wen, “New Robust Face Recognition Methods Based on Linear Regression,” Plos One, vol. 7, no. 8, pp. e42461, 2012.

[5] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 2, pp. 210-227, Feb, 2009.

[6] Y. Meng, Z. Lei, Y. Jian, and D. Zhang, “Regularized Robust Coding for Face Recognition,” Image Processing, IEEE Transactions on, vol. 22, no. 5, pp. 1753-1766, 2013.

[7] R. He, W.-S. Zheng, and B.-G. Hu, “Maximum Correntropy Criterion for Robust Face Recognition,” Ieee Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 8, pp. 1561-1576, Aug, 2011.

[8] H. Ran, Z. Wei-Shi, T. Tieniu, and S. Zhenan, “Half-Quadratic-Based Iterative Minimization for Robust Sparse Representation,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 36, no. 2, pp. 261-275, 2014.

[9] I. Naseem, R. Togneri, and M. Bennamoun, “Linear Regression for Face Recognition,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 11, pp. 2106-2112, 2010.

[10] D. Zhang, Y. Meng, and F. Xiangchu, "Sparse representation or collaborative representation: Which helps face recognition?." pp. 471-478.

[11] 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, Sep, 2011.

[12] J.-X. Mi, “Face image recognition via collaborative representation on selected training samples,” Optik-International Journal for Light and Electron Optics, vol. 124, no. 18, pp. 3310-3313, 2013.

[13] Y. Liu, J.-X. Mi, C. Li, and C. Li, “Competitive Sparse Representation Classification for Face Recognition,” International Journal of Advanced Computer Science & Applications, vol. 1, no. 6, pp. 1-7.

[14] A. M. Martinez, “The AR face database,” CVC Technical Report, vol. 24, 1998.

[15] F. S. Samaria, and A. C. Harter, "Parameterisation of a stochastic model for human face identification." pp. 138-142.

[16] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, “The FERET database and evaluation procedure for face-recognition algorithms,” Image and vision computing, vol. 16, no. 5, pp. 295-306, 1998.

[17] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[18] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, “From few to many: Illumination cone models for face recognition under variable lighting and pose,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 23, no. 6, pp. 643-660, 2001.

[19] K. C. Lee, J. Ho, and D. J. Kriegman, “Acquiring linear subspaces for face recognition under variable lighting,” IEEE Trans Pattern Anal Mach Intell, vol. 27, no. 5, pp. 684-98, May, 2005.