

## Research Article

# Integrating Globality and Locality for Robust Representation Based Classification

Zheng Zhang,<sup>1,2</sup> Zhengming Li,<sup>1,3</sup> Binglei Xie,<sup>2</sup> Long Wang,<sup>2</sup> and Yan Chen<sup>2,4</sup>

<sup>1</sup> Key Laboratory of Network Oriented Intelligent Computation, Shenzhen, Guangdong 518055, China

<sup>2</sup> Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, Guangdong 518055, China

<sup>3</sup> Guangdong Industry Training Center, Guangdong Polytechnic Normal University, Guangzhou, Guangdong 510665, China

<sup>4</sup> Shenzhen Sunwin Intelligent Corporation, Shenzhen, Guangdong 518055, China

Correspondence should be addressed to Zheng Zhang; [darrenzz219@gmail.com](mailto:darrenzz219@gmail.com)

Received 24 December 2013; Accepted 21 February 2014; Published 26 March 2014

Academic Editor: Carsten Proppe

Copyright © 2014 Zheng Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The representation based classification method (RBCM) has shown huge potential for face recognition since it first emerged. Linear regression classification (LRC) method and collaborative representation classification (CRC) method are two well-known RBCMs. LRC and CRC exploit training samples of each class and all the training samples to represent the testing sample, respectively, and subsequently conduct classification on the basis of the representation residual. LRC method can be viewed as a “locality representation” method because it just uses the training samples of each class to represent the testing sample and it cannot embody the effectiveness of the “globality representation.” On the contrary, it seems that CRC method cannot own the benefit of locality of the general RBCM. Thus we propose to integrate CRC and LRC to perform more robust representation based classification. The experimental results on benchmark face databases substantially demonstrate that the proposed method achieves high classification accuracy.

## 1. Introduction

Face recognition has become a popular technique as one of the most promising branches of the pattern recognition [1–4]. Recently, the representation based classification method (RBCM) was proposed for face recognition by the specialists and scholars in the fields of computer vision and pattern recognition [5–7]. The RBCM are completely different from the conventional classification methods, such as principle component analysis (PCA) [8] and kernel PCA [9], linear discriminant analysis (LDA) [10], kernel LDA [11], and Gaussian maximum likelihood [12]. The common procedure of these traditional methods is to first utilize all the training samples to obtain the transform axes and then exploit the predetermined transform axes to convert all the samples into a new lower-dimensional space to perform recognition [4, 8–12]. However, the RBCM generally uses the linear combination of the training samples to represent the test sample and then utilizes the representation results to classify the test sample. The RBCM has been widely applied to the face

recognition [13–15], image categorization [16, 17], and image superresolution [18, 19].

Generally speaking, the representation based classification consists of two categories, the global representation method and the local representation method. Hereafter, the global representation method refers to use all the training samples to represent the test sample, whereas the local representation method just uses the fraction of the training samples to represent the test sample. What is more, from the viewpoint of the image restoration, the linear combination of training samples can be viewed as a new image with the same size of the samples. If the linear combination of training samples from a class can perfectly reconstruct the probe image, the probe image can be classified as a member of this class. The corresponding residual can be regarded as the distance between the test sample and each class. Both of the global and local representation methods make use of the residual to perform final classification.

Since the sparse representation based classification (SRC) method [5] was proposed, the sparse representation based

classification has attracted tremendous attention and plenty of representation based classification methods were designed to improve the effectiveness and efficiency of SRC [1, 2, 20, 21]. SRC can be somewhat viewed as the global representation method and SRC exploits the linear combination of all the training samples with a sparse optimizer to represent the given test sample. Moreover, the sparse optimizer is usually implemented by the  $l_0$ -norm,  $l_p$ -norm ( $0 < p < 1$ ),  $l_1$ -norm, and even  $l_2$ -norm minimization. Pati et al. [22] proposed the orthogonal matching pursuit (OMP) method exploiting the  $l_0$ -norm minimization to obtain the sparse representation solution. Xu et al. [23] exploited the  $l_{1/2}$ -norm minimization to restrain the sparsity of representation coefficients. The RBCM with  $l_1$ -norm or  $l_2$ -norm minimization constraint seems to be the mainstream approaches [24–26]. The literature [27–29] has proposed some powerful  $l_1$ -norm and  $l_2$ -norm minimization based algorithms for SRC. For example, Beck and Teboulle [27] proposed a fast iterative shrinkage-thresholding algorithm (FISTA) using Nesterov's optimal gradient-based algorithm for convex optimization problems. Kim et al. [29] proposed a novel method exploiting conjugate gradients algorithm to address the convex quadratic problem. However, SRC suffers from distinct time complexity on account of solving the minimization problem iteratively. Thus abundant research communities convert their attention to develop RBCM with loose sparse restriction,  $l_2$ -norm minimization constraint. We briefly review some remarkable literature resources in the following. Zhang et al. [30] presented an efficient representation method called collaborative representation classification (CRC) method. CRC method introduces  $l_2$ -norm minimization, rather than  $l_1$ -norm minimization, to acquire limited-sparse representation solution. Furthermore, it is confirmed that the computational efficiency of CRC is extremely higher than SRC on the basis of analogical classification accuracy [30]. Xu et al. [4] put forward a two-step sparse representation method which first utilized CRC method to figure out  $M$  nearest samples and then implemented CRC on the predetermined  $M$  samples to make final classification. Liu et al. [31] used reconstruction error and normalized distance of RBCM to perform palmprint recognition method. Xu et al. [32] applied the model of RBCM to improve the nearest neighbor classifier. Yang et al. [26] presented a relaxed collaborative representation classification using the similar and distinctive features. Xu et al. [33] proposed to utilize the bimodal biometric traits to perform sparse representation method for biometrics and palmprint classification. The literature [25, 26, 30–34] also certifies the robustness and efficiency of CRC method in contrast to SRC. Furthermore, CRC also endows with the characteristic of the limited sparsity [30, 34].

Local recognition methods have been proposed to improve the classification accuracy and have extraordinary priority in comparison with the global recognition methods [35–37]. The local recognition methods only exploit part of the training samples rather than the whole of training samples to obtain intermediate result and then use the result to make classification. For example, Xu et al. [38] proposed a novel and efficient solution scheme for locality preserving projection (LPP) method and this method can effectively conquer the

small training sample size problem. Sugiyama [39] proposed a local LDA method that combined the advantages of LDA and the preponderance of LPP method to perform supervised dimensionality reduction. Liu et al. [40] demonstrated that the local principle component analysis (PCA) had great superiorities compared to the global PCA. In order to obtain better face recognition results, local RBCM method recently was proposed to exploit each class of the training samples to represent the test sample and classify the test sample based on the representation residual. Linear regression classification (LRC) method was proposed to utilize each class of the training samples to represent the test sample and classify the test sample into the class which generates the minimum representation residual [35]. Naseem et al. [36] developed a robust regression method using a linear combination of specific class of training sample to represent the test sample. Tahir et al. [37] fused the histogram and face features using LRC classifier to constitute a texture descriptor, named multiscale local phase quantization histogram (MLPQH) feature, to solve the illumination problem for face recognition. The literature [35–37] certified the significance of locality in RBCM and also demonstrated that the local representation methods could greatly lower the recognition error rate.

In order to obtain the bilateral dominant characteristics of the global and local RBCM, we integrate CRC and LRC together to perform face recognition. CRC method possesses the property of limited sparsity and can properly harness the nature of low-complex sparsity to perform global representation based classification. The LRC method is a local representation method and is able to employ the local structure of patterns to perform classification. Section 4 verifies the rationality of the proposed method and the substantial experiments in Section 5 also confirm its effectiveness.

The rest of this paper is organized as follows: we first briefly review the LRC and CRC methods in Section 2. In Section 3, we present the proposed optimizing method. In Section 4, we discuss the potential rationale of the proposed method. In Section 5, the proposed method is verified by conducting extensive experiments on several well-known benchmark face databases. Finally, Section 6 concludes the paper.

## 2. Related Work

**2.1. Linear Regression Classification Algorithm.** Suppose that there are  $c$  known pattern classes with  $N$  training samples  $x_1, x_2, \dots, x_N$  and the query sample is  $y$ . All the samples should be firstly normalized to unit column vectors with the length of 1.

Let  $X_i$  be the matrix composed by the training samples of the  $i$ th class; that is,  $X_i = [x_{i1}, x_{i2}, \dots, x_{iM_i}] \in R^{d \times M_i}$  where  $d$  is the dimension of the sample and  $M_i$  is the number of training samples of the  $i$ th class. Let us note a matrix  $X = [X_1, X_2, \dots, X_c] \in R^{d \times N}$  and, apparently, the matrix  $X$  denotes the whole of training samples.

Linear regression classification (LRC) method [35] first exploits a linear combination of each class of training samples

to represent the testing sample  $y$ ; that is, the following equation is satisfied:

$$y = a_1 x_{i1} + a_2 x_{i2} + \cdots + a_n x_{iM_i}, \quad (1)$$

where  $x_{i1}, x_{i2}, \dots, x_{iM_i}$  are the  $i$ th class of the training samples and  $a_j$  ( $j = 1, 2, \dots, n$ ) is the coefficient of  $x_{ij}$ . For convenient presentation, we can rewrite (1) into the following equation:

$$y = X_i \beta_i, \quad (2)$$

where  $X_i = [x_{i1}, x_{i2}, \dots, x_{iM_i}] \in R^{d \times M_i}$ ,  $\beta_i = [a_1, a_2, \dots, a_n]^T$ , and  $y \in R^{d \times 1}$  are column vectors.

According to the least square theory,  $\beta$  can be calculated by the equation below:

$$\beta_i = (X_i^T X_i)^{-1} X_i^T y, \quad i = 1, 2, \dots, c. \quad (3)$$

We refer to  $\tilde{y} = X_i \beta_i$  as the representation result of the  $i$ th class of the training samples using preobtained  $\beta_i$  and we can transform  $\tilde{y}$  into a two-dimension image with the same size as the original sample image.

Finally, we calculate the representation residuals of each class of the training samples and then exploit the representation residuals to conduct classification. The residual of representing the testing sample,  $y$ , using the training samples of the  $i$ th class is

$$\text{resi}(i) = \|y - X_i \beta_i\|^2. \quad (4)$$

We note  $X_i \beta_i$  as the representation contribution using the training samples from the  $i$ th class and the RBCM presumes that the test sample,  $y$ , can be adequately represented by only the training samples from the same class [41]. Thus, a smaller representation residual can reflect a larger representation contribution, and then the minimum representation residual,  $\text{resi}(i)$ , means the nearest distance between the testing sample and the  $i$ th class. Consequently, we classify the testing sample,  $y$ , into the class which achieves the smallest representation residual.

**2.2. Collaborative Representation Classification Algorithm.** Collaborative representation classification (CRC) [30] method first utilizes all the training samples to represent the test sample and exploits the linear combination of the whole of training samples to approximate the test sample. That is, the following approximation is satisfied:

$$y = a_1 x_1 + a_2 x_2 + \cdots + a_N x_N \quad (5)$$

or

$$y = XA, \quad (6)$$

where  $y$  is the testing sample. The  $a_i$  ( $i = 1, 2, \dots, N$ ) is the coefficient of  $x_i$  and  $A = [a_1 a_2 \cdots a_N]^T$ . The solution of  $A$  can be calculated by solving  $l_2$ -norm minimization problem:

$$\widehat{A} = \arg \min \|A\|_2 \quad \text{s.t. } y = XA \quad (7)$$

or

$$\min (\|y - XA\|_2 + \mu \|A\|_2), \quad (8)$$

where  $\mu$  is a small positive constant and (8) is to utilize the Lagrangian Method to obtain the solution of (7). We solve (8) by  $\widehat{A} = (X^T X + \mu I)^{-1} X^T y$ , where  $I$  is the identity matrix.

Similarly, we evaluate the representation contributions and representation residuals of each class of the training samples using

$$\text{devi}(i) = \left\| y - \sum_{k=(i-1) \times m+1}^{i \times m} X_k \widetilde{A}_k \right\|_2, \quad (9)$$

where  $\sum_{k=(i-1) \times m+1}^{i \times m} X_k \widetilde{A}_k$  and  $\text{devi}(i)$  are viewed as the representation contribution and the representation residuals of the training samples from the  $i$ th class, respectively, and then classify the testing sample,  $y$ , into the class which leads to the minimum representation residual.

### 3. The Proposed Method

This section describes the proposed method of integrating CRC and LRC in detail. As presented above, (1) just uses training samples of each class to represent the test sample, whereas (5) uses all the training samples to represent the test sample.

The literature [30, 34] has proved that CRC method possesses limited-sparse representation capability to perform an excellent RBCM for face recognition. The literature [35, 36] also has shown the LRC method owns the ability to address the challenges of varying facial expressions, illumination, and occlusion. Thus, the motivation of the proposed method is to combine LRC and CRC to make robust and effective face recognition.

Therefore, we propose a novel model by integrating CRC method and LRC method. Then the objective function can be constructed by

$$\arg \min (\|y - XA\|_2 + \mu \|A\|_2 + \gamma \|y - X_i A_i\|), \quad (10)$$

where  $y$  is the test sample and  $X = [X_1 \cdots X_c]$ .  $X_i$  is the matrix consisting of the training samples from the  $i$ th class.  $A$  is the coefficient of  $X$  and  $A_i$  is the coefficient of  $X_i$ .  $\gamma$  and  $\mu$  are small positive constants.

Equation (10) is a joint optimization problem of CRC in (8) and LRC in (2). The main steps of the proposed method to solve (10) are described as follows.

*Step 1.* Normalizing all the samples including the training samples and test samples to unit vectors of length 1 using  $x_i = x_i / \|x_i\|$  and  $y_i = y_i / \|y_i\|$ .

*Step 2.* Using CRC method by solving (8) to obtain each class of representation residual, that is, applying (9) to compute the representation residual of the  $j$ th class,  $\text{devi}(j)$ .

*Step 3.* Using LRC method by solving (2) to obtain each class of the representation residual, that is, utilizing (4) to calculate the representation residual of the  $i$ th class,  $\text{resi}(i)$ .

*Step 4.* Normalizing representation residuals, devi and resi. For test sample  $y$ , devi and resi are normalized by

$$\text{devi}' = \frac{\text{revi} - \text{revi}_{\min}}{\text{devi}_{\max} - \text{devi}_{\min}}, \quad \text{resi}' = \frac{\text{resi} - \text{resi}_{\min}}{\text{resi}_{\max} - \text{resi}_{\min}}, \quad (11)$$

where  $\text{devi}_{\max}$  and  $\text{resi}_{\max}$  denote the maximum representation residual of devi and resi, respectively.  $\text{devi}_{\min}$  and  $\text{resi}_{\min}$  are the minimum representation residual of devi and resi.

*Step 5.* Integrating two different residuals obtained by CRC and LRC. For the test sample  $y$ , we calculate the fusing score of the test sample with respect to the training samples of  $i$ th class using

$$\text{dist}_i = \text{devi}'_i + \gamma \text{resi}'_i, \quad (12)$$

where  $\gamma$  is the weight of the representation residual.

*Step 6.* Outputting the identity of the test sample  $y$  as

$$\text{identity}(y) = \arg \min_i (\text{dist}). \quad (13)$$

We believe that a lower representation residual can lead to a more accurate representation and we assign the test sample to the class with the smallest final representation residual.

#### 4. Potential Rationale of the Proposed Method

In this section, we analyze our method for exploring its underlying characteristics. We will discuss the effectiveness of our method from the following three aspects: the significance of the proposed method, the advantages of the local representation based classification, and the superiorities of the global representation based classification.

CRC method and LRC method use different ideas of ideologies to make representation based classification and the correlation coefficients can reflect the difference between the two methods. Figure 1 illustrates the correlation coefficients of these two methods. We treated the first three samples of each subject in ORL database [42] as the training samples and regarded the rest of the samples as test samples and then calculated the representation residual of each class using LRC and CRC, respectively. And the representation residuals from all the classes for each testing sample can form a residual vector, and then we compute the correlation coefficients of residual vectors obtained by CRC and LRC. From Figure 1, we can see that the correlation coefficients are from 0.2148 to 0.7765. That means CRC and LRC have weak relationships with each other. Thus, combining the two methods is meaningful and significant. For example, we conducted an experiment on ORL database and used the first sample from each subject as training sample and the rest of the samples as test samples. The testing sample shown in Figure 2(a) from the 21th class is classified into the 36th class, Figure 2(b), using CRC, but LRC assigns the testing sample to the 19th class, Figure 2(c). However, our method can make a correct classification into the 21th class, Figure 2(d). Figure 3 presents the representation residuals

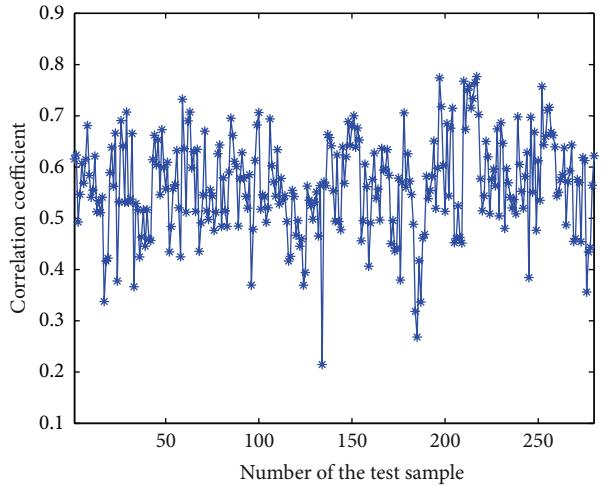


FIGURE 1: The correlation coefficient of CRC and LRC. We exploit the first three samples of each subject as training sample on the ORL database and the remaining samples as test samples. And the figure illustrates the correlation coefficient of representation residuals of each testing sample using CRC and LRC.

using CRC, LRC, and the proposed method, respectively. Moreover, the experimental results in Section 5 also verify that it is practical to exploit the proposed method to make effective classification.

Representation based classification method (RBCM) usually supposes that only the training samples from the same class can sufficiently represent the test sample [41]. However, if one subject extremely resembles another subject, the above assumption may not stand. For example, we performed an experiment on FERET face database [43] using the first sample of each subject as training sample and the remaining samples for testing samples. Figure 4 shows that for the test sample wrongly and accurately classified by CRC and LRC method, our method certainly can assign the test sample to the right class. Figure 5 presents the coefficients of the representation solution, and it is apparent that the largest two coefficients are the coefficients of training samples from the 110th class and the 72th class. However, CRC assigns the test sample to the 72th class that is the nearest class. On the contrary, LRC and our method can exactly recognize the test sample. Thus, the local representation based classification method can compensate for the imperfection of the global representation based recognition method. From the viewpoint of the proposed method, our method can sufficiently implement the two typical characteristics of the local representation based classification, LRC, and the global representation based classification method, CRC.

It is obvious that global representation based classification method, which uses all the training samples to represent the test sample, can alleviate the difficulty of the small sample size problem. For example, if we just use one sample of each subject to make classification by exploiting LRC method, LRC will deteriorate to the nearest neighbor (NN) classifier and also lose the preponderance of RBCM. In that regard, score

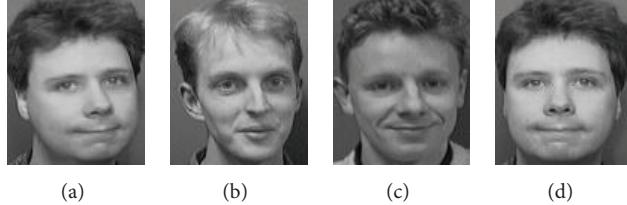


FIGURE 2: A test sample that is erroneously and correctly classified by CRC, LRC, and the proposed method, respectively. (a) The test sample. (b), (c) One sample from the class wrongly classified by CRC and LRC. (d) One sample from the class correctly classified by our method.

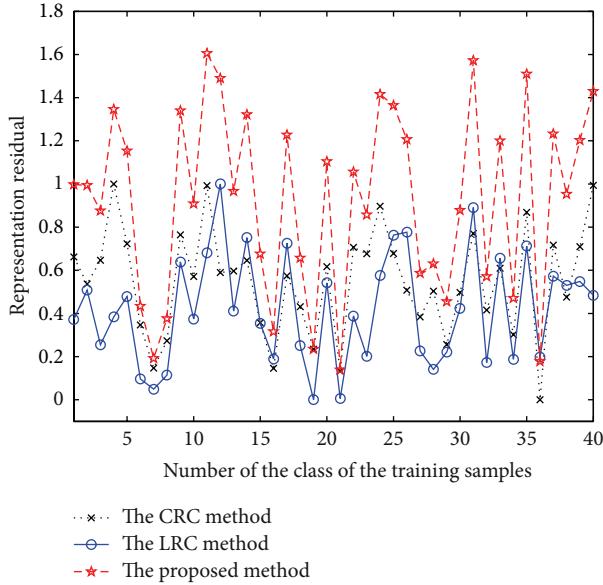


FIGURE 3: The representation residuals of the test sample obtained by CRC, LRC, and our method. We use the first sample of each subject as training sample and the remaining samples as test samples on the ORL database. We can see that the minimum representation residuals using CRC and LRC are the 36th and 19th class of training samples, respectively, but the smallest representation residual using our method is the 21th training sample.

fusion of CRC and LRC is meaningful and can obtain better result.

CRC and SRC are the typical examples of global RBCMs and LRC is one of the typical local RBCMs. However, SRC has very high computation complexity and CRC owns the superiority that CRC is extremely faster than SRC with very competitive face recognition accuracy. What is more, CRC and LRC will be almost not able to identify the class to which the test sample truly belongs and will erroneously classify the test sample. However, the proposed method can compensate for the imperfection and shortcomings of each other. Moreover, the correlation coefficients between these two methods are very low. Thus, integrating CRC and LRC can take full advantages of these two methods. At this stage, we can demonstrate that the framework of integrating the advantages of global RBCM and the superiorities of local RBCM is reasonable and meaningful.

TABLE 1: Rates of classification errors of different methods on the ORL face database.

	1	2	3	4	5
Our method ( $\gamma = 0.9$ )	0.2944	0.1500	0.1321	0.0833	0.1100
Our method ( $\gamma = 0.85$ )	0.2972	0.1469	0.1357	0.0833	0.1100
Our method ( $\gamma = 0.8$ )	0.2972	0.1500	0.1321	0.0833	0.1050
CRC	0.3194	0.1656	0.1392	0.1083	0.1150
LRC	0.3250	0.2063	0.1857	0.1375	0.1200
INNC	0.3194	0.2125	0.2178	0.1750	0.1700
LDA	—	0.2531	0.1678	0.1458	0.1750
SRC	0.3177	0.1781	0.1607	0.1250	0.1200

## 5. Experimental Results

In this section, we conducted extensive experiments to evaluate the effectiveness and robustness of the proposed method on different face databases including ORL, AR, CMU PIE, and FERET database. In order to present the performance of the proposed method, several competitive face recognition methods are tested as comparison, such as CRC [30], LRC [35], LDA, improved nearest neighbor classification (INNC) method [32], and the sparse representation based classification (SRC) algorithm proposed in [29]. Moreover, we also performed experiments on corrupted face databases and conducted contiguous occlusion experiments. The parameter of the proposed method is set to  $\mu = 0.01$  for all the experiments.

**5.1. Experiment on the ORL Face Database.** The ORL face database [42] includes 400 face images obtained from 40 distinct subjects and each subject provides ten different face images. For some subjects, the images were taken at different times, varying the lighting conditions, facial expressions (open or closed eyes, smiling, or not smiling), and facial details (glasses or no glasses). All the images were captured against a dark homogeneous background with the subjects in an upright, frontal position. In order to reduce the computational complexity, all the images were resized to 56 by 46 image matrix at first. Figure 6 presents some images from ORL database. When the number of training sample is one, LDA algorithm is not available since the between-class scatter matrix is zero.

We took the first 1, 2, 3, 4, and 5 image samples of each subject as original training samples, respectively, and regarded the remaining face images as test samples. Table 1



FIGURE 4: A test sample that is falsely and accurately classified by CRC and LRC. (a) A test sample from the 110th class. (b) A sample from the 72th class which is the closest to the test sample. (c) One sample from the 110th class precisely classified by LRC method and the proposed method.

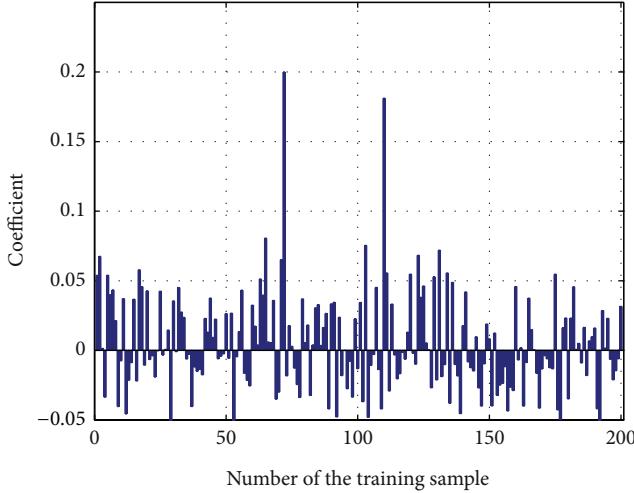


FIGURE 5: The coefficients of the training samples obtained by CRC. The first sample of each subject on the FERET database is regarded as training sample and the remaining samples are treated as test samples. This figure illustrates the coefficient of training samples exploiting all the training samples to represent the testing sample, Figure 4(a), from the FERET face database.

enumerates the variation of the classification error rates of different solutions. And our method with different values of  $\gamma$  also verifies that the proposed method outperforms the other classification methods. For example, when we took the first four face images of each subject as training samples and the rest of the face images as test samples, the classification error rate of our method can greatly outperforms the CRC, LRC, INNC, LDA, and SRC algorithms by a margin of 2.5 percent, 5.42 percent, 9.17 percent, 6.25 percent, and 4.17 percent, respectively. What is more, it is clear that the proposed method can dramatically improve CRC and LRC classification methods and it also reflects the effectiveness of the framework of integrating the global RBCM and local RBCM.

**5.2. Experiments on the AR Face Database.** The AR face database [44] is composed of more than 3000 face images of 126 subjects and 26 facial images for each subject were taken in two separated terms and the resolution of each AR image has been resized to a 40 by 50 image matrix. This face database has much more variations, such as, illumination conditions, facial expressions, and facial disguises. We

TABLE 2: Rates of classification errors of different methods on the AR face database.

	1	2	3	4
Our method ( $\gamma = 0.90$ )	0.2906	0.3045	0.3061	0.3166
Our method ( $\gamma = 0.75$ )	0.2890	0.3059	0.3029	0.3125
Our method ( $\gamma = 0.65$ )	0.2887	0.3069	0.2996	0.3140
CRC	0.3017	0.3167	0.3141	0.3246
LRC	0.3807	0.4038	0.4094	0.4186
INNC	0.3017	0.3299	0.3301	0.3413
LDA	—	0.3965	0.3880	0.4102
SRC	0.3120	0.3156	0.3167	0.3220

treated the first  $1, 2, \dots, 9$  face images from each subject as the original training samples and regarded the remaining samples as test samples, respectively. Figure 7 has shown some face images for four subjects from the AR database. Table 2 shows the classification error rates using different classification methods. It is obvious that our method has remarkable priorities compared to the other classification algorithms. For example, when we took the first face image of each subject as training samples and the rest of the face images as test samples, the classification error rates of our method with  $\gamma = 0.90$ ,  $\gamma = 0.75$ ,  $\gamma = 0.65$ , CRC, LRC, INNC, and SRC are 29.06%, 28.90%, 28.87%, 30.17%, 38.07%, 30.17%, and 31.20%, respectively. That means that 9.2 percent recognition errors can be avoided by using our method instead of LRC. We can see that combining CRC and LRC is a favorable idea to perform face recognition.

**5.3. Experiments on FERET Face Database.** The FERET database is one of the largest publicly available facial images databases. We just selected a subset made up of 1400 images from 200 individuals and each one contains seven images to conduct experiments [43]. The images were collected in a semicontrolled environment. We resized each image to a 40 by 40 image matrix using the same downsampling algorithm. For some individuals, there were over two years between their first and last time to be photographed. Figure 8 presents some cropped face images from the FERET face database. The parameter of our method is set to  $\gamma = 1.0$  for this experiment.

We perform an experiment on FERET face database using the first six samples of each subject as training sample and the remaining samples for testing. The classification error rates exploiting our method, CRC, LRC, INNC, LDA, and SRC are 15.00%, 25.50%, 21.50%, 25.50%, 41.50%, and 18.50%,

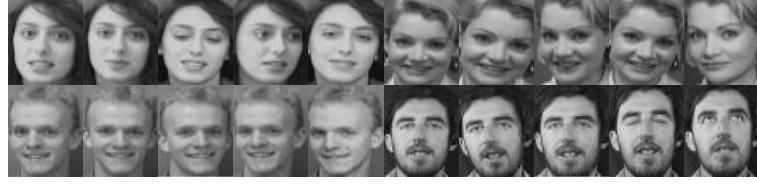


FIGURE 6: Some face images from the ORL face database.



FIGURE 7: Some face images from the AR face database.



FIGURE 8: Some face images from the FERET face database.

respectively. In other words, this suggests that the proposed method can greatly outperform the CRC, LRC INNC, LDA, and SRC algorithms by a margin of 10.5 percent, 6.5 percent, 10.5 percent, 26.5 percent, and 2.5 percent, respectively.

The first image of each subject was used for training sample, while the remaining six were served as test samples. Also Figure 9 shows the detailed experimental results compared to other classification methods for a large variation of feature dimensions. From Figure 9, we can see that our method achieves the lowest classification error rate compared to the other classification methods although the dimensions of the samples are extremely low. The classification error rates under different feature dimensions suggest that the proposed method is superior to other classifiers.

**5.4. Experiments on CMU PIE Database.** We chose a subset of the CMU PIE database to perform experiments. For each subject in this face database, we randomly selected 21 face images under different lighting conditions and expressions including 1428 images of 68 subjects for experiments [45]. The size of each image is cropped to 32 by 32 pixels. We just took the first 1, 2, and 3 face images of each subject as the training samples and the rest of the samples as test sample. The experimental results are shown in Table 3 and we can draw a conclusion that the proposed method can achieve the

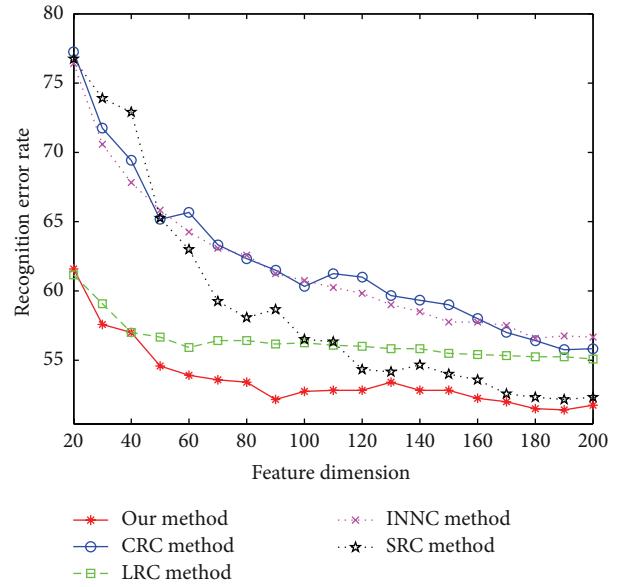


FIGURE 9: The recognition error rates of different method under different feature dimensions. We treat the first image of each subject as training sample from FERET face database and the remaining images are used for test samples. The figure shows the detailed experimental results compared to other classification methods for a large variation of feature dimensions.



FIGURE 10: Some corrupted face images in ORL database and AR database. The first two rows show the corrupted face images used in the ORL database. The last two rows show the corrupted face images used in the AR face database.

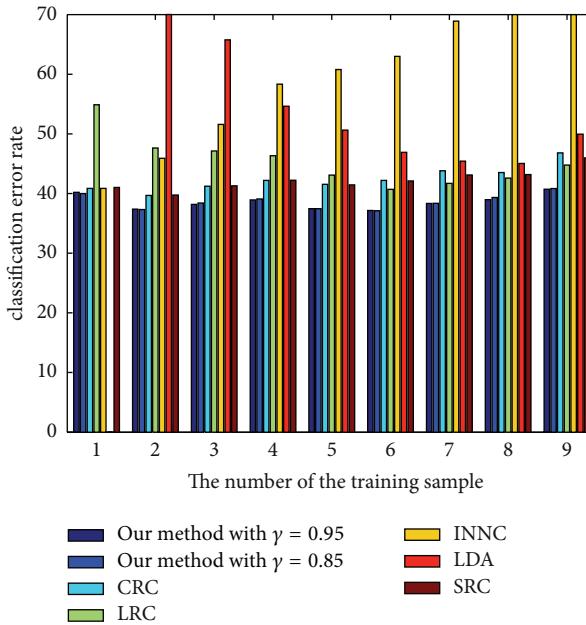


FIGURE 11: Rates of classification error rates of different methods on the corrupted AR face database. We exploit the first  $1, 2, \dots, 9$  face images of each subject as training samples and take the remaining samples as test samples. The figure shows the recognition error rates using our method with the parameter  $\gamma = 0.95$  and  $\gamma = 0.85$ , CRC, LRC, INNC, LDA, and SRC, respectively.

minimum classification error rate. For example, when the number of training sample is one, the classification error rates using our method with  $\gamma = 1.0$ ,  $\gamma = 0.85$ ,  $\gamma = 0.70$ , CRC, LRC, INNC, and SRC are 10.80%, 11.17%, 11.32%, 16.10%, 13.82%, 16.10%, and 16.10%, respectively. Thus the proposed method can extensively improve the classification accuracy compared to CRC and LRC methods and our method obtains the lowest classification error rates.

**5.5. Recognition on Random Pixel Corruption Databases.** We performed experiments on random pixel corrupted face images to test the robustness of the methods. The datasets of

TABLE 3: Rates of classification errors of different methods on the illumination variation subset of CMU PIE face database.

	1	2	3
The proposed method ( $\gamma = 1.0$ )	0.1080	0.0542	0.0049
The proposed method ( $\gamma = 0.85$ )	0.1117	0.0503	0.0049
The proposed method ( $\gamma = 0.70$ )	0.1132	0.0479	0.0032
CRC	0.1610	0.0975	0.0080
LRC	0.1382	0.1060	0.0849
INNC	0.1610	0.0588	0.0049
LDA	—	0.1262	0.0131
SRC	0.1610	0.1370	0.0179

TABLE 4: Rates of classification errors of different methods on corrupted ORL face database.

	2	3	4	5
Our method ( $\gamma = 0.65$ )	0.1656	0.1500	0.0916	0.0900
Our method ( $\gamma = 0.60$ )	0.1656	0.1500	0.0875	0.0950
CRC	0.1844	0.1714	0.1083	0.1100
LRC	0.2594	0.2286	0.1625	0.1450
INNC	0.2312	0.2178	0.1583	0.1600
LDA	0.4625	0.2964	0.2125	0.1600
SRC	0.1813	0.1679	0.1083	0.1000

training samples and the test samples are the same as those in Sections 5.1 and 5.2, respectively. In order to obtain random pixel corruption face images, we add Gaussian white noise of zero mean and variance of 0.01 to the ORL database and SaltPepper noise of density of 0.1 to AR database, respectively. Figure 10 shows some corrupted face images on ORL database and AR database. Table 4 and Figure 11 show the classification error rate results on ORL database and AR database, respectively. The LDA algorithm is not available when the number of the training sample for each class is one, so we set the classification error rate exploiting the LDA algorithm to 0 for corrupted AR database in Figure 11. We see again that our method possesses the absolute dominance compared to all the other methods. For example, when each subject provides 3 corrupted ORL face images as training samples,

the classification error rates using our method with  $\gamma = 0.65$ , CRC, LRC, INNC, LDA, and SRC are 15.00%, 17.14%, 22.86%, 21.78%, 29.64%, and 16.79%, respectively. Thus the experimental results also demonstrate the robustness and effectiveness of the proposed method.

## 6. Conclusion

The proposed method combines CRC and LRC methods to perform effective and robust face recognition. The significance of our method is to sufficiently exploit the globality of CRC and the locality of LRC, respectively. The potential rationale of this method is that integrating LRC and CRC is meaningful because they have the low correlation coefficients showed in Figure 1. Moreover, it is remarkable that our method obtains a promising performance on face databases with random pixel corruption, and then we can confirm the good efficiency of the proposed method under complex conditions.

All the experiments in Section 5 can testify that the framework of integrating the globality and locality of training samples can tremendously improve the classification accuracy. CRC, as one of the typical global RBCMs, can completely utilize the globality of all the training samples, whereas LRC, as one of the typical local RBCMs, can make the most of the locality of the training samples. A large number of experiments also demonstrate that the proposed framework is reasonable and practical and it can extensively improve the classification accuracy compared to the state-of-art classification methods on different face databases. Furthermore, the proposed method is simple and computationally efficient. Besides the method description and the experimental analysis, the rationale of our method also has been explicitly interpreted and the superiority of our method also has been visually presented in this paper.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This paper is partially supported by the National Natural Science Foundation of China (Grant nos. 61300032 and 61370163) and Shenzhen Municipal Science and Technology Innovation Council (nos. JCYJ20130401152508661 and JCYJ20130329151843309).

## References

- [1] R. He, W. S. Zheng, T. Tan, and Z. Sun, “Half-quadratic based iterative minimization for robust sparse representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 261–275, 2014.
- [2] W. Deng, J. Hu, and J. Guo, “Extended SRC: undersampled face recognition via intraclass variant dictionary,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1864–1870, 2012.
- [3] S.-Y. Cho, C. W. Ting, and C. Quek, “Thermal facial pattern recognition for personal verification using fuzzy cmac model,” *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 1, pp. 203–222, 2011.
- [4] 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.
- [5] 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, no. 2, pp. 210–227, 2009.
- [6] Y. Xu, X. Zhu, Z. Li, G. Liu et al., “Using the original and ‘symmetrical face’ training samples to perform representation based two-step face recognition,” *Pattern Recognit*, vol. 46, no. 4, pp. 1151–1158, 2013.
- [7] S. Ergin and M. B. Gulmezoglu, “A novel framework for partition-based face recognition,” *International Journal of Innovative Computing Information and Control*, vol. 9, no. 5, pp. 1819–1834, 2013.
- [8] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [9] M. Debruyne and T. Verdonck, “Robust kernel principal component analysis and classification,” *Advances in Data Analysis and Classification*, vol. 4, no. 2–3, pp. 151–167, 2010.
- [10] S. W. Park and M. Savvides, “A multifactor extension of linear discriminant analysis for face recognition under varying pose and illumination,” *Eurasip Journal on Advances in Signal Processing*, vol. 2010, Article ID 158395, 2010.
- [11] K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, “An introduction to kernel-based learning algorithms,” *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–201, 2001.
- [12] C.-P. Lu and J.-J. Liaw, “Yarn tension pattern retrieval system based on Gaussian maximum likelihood,” *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 11, pp. 6261–6272, 2011.
- [13] P. Nagesh and B. Li, “A compressive sensing approach for Expression-invariant face recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1518–1525, June 2009.
- [14] Y. Xu, Q. Zhu, and D. Zhang, “Combine crossing matching scores with conventional matching scores for bimodal biometrics and face and palmprint recognition experiments,” *Neurocomputing*, vol. 74, no. 18, pp. 3946–3952, 2011.
- [15] A. Yang, Z. Zhou, A. Ganesh et al., “Fast l1-minimization algorithms for robust face recognition,” *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3234–3246, 2013.
- [16] J. Mairal, F. Bach, J. Ponce et al., “Discriminative dictionaries for local image analysis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’08)*, pp. 1–8, 2008.
- [17] S. Gao, I. W.-H. Tsang, and L.-T. Chia, “Kernel sparse representation for image classification and face recognition,” in *Proceedings of the Computer Vision (ECCV ’10)*, vol. 6314, pp. 1–14, 2010.
- [18] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’08)*, pp. 1–8, June 2008.
- [19] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

- [20] W. Deng, J. Hu, and J. Guo, "In defense of sparsity based face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern*, pp. 399–406, 2013.
- [21] R. He, W. Zheng, and B. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 1561–1576, 2011.
- [22] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of the IEEE 27th Asilomar Conference on Signals, Systems & Computers*, pp. 40–44, November 1993.
- [23] Z. Xu, H. Zhang, Y. Wang, X. Chang, and Y. Liang, " $L_{1/2}$  regularization," *Science China. Information Sciences*, vol. 53, no. 6, pp. 1159–1169, 2010.
- [24] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.
- [25] 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.
- [26] M. Yang, L. Zhang, D. Zhang et al., "Relaxed collaborative representation for pattern classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2224–2231, 2012.
- [27] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [28] M. S. Asif and J. Romberg, "Sparse recovery of streaming signals using L1-homotopy," *IEEE Transactions on Signal Processing*. In press.
- [29] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale  $\ell_1$ -regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.
- [30] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 471–478, November 2011.
- [31] Z. Liu, J. Pu, T. Huang, and Y. Qiu, "A novel classification method for palmprint recognition based on reconstruction error and normalized distance," *Applied Intelligence*, vol. 39, no. 2, pp. 307–314, 2013.
- [32] Y. Xu, Q. Zhu, Y. Chen et al., "An improvement to the nearest neighbor classifier and face recognition experiments," *International Journal of Innovative Computing Information and Control*, vol. 9, no. 2, pp. 543–554, 2013.
- [33] Y. Xu, Q. Zhu, Z. Fan et al., "Coarse to fine K nearest neighbor classifier," *Pattern Recognition Letters*, vol. 34, no. 9, pp. 980–986, 2013.
- [34] L. Zhang, M. Yang, X. Feng et al., "Collaborative representationbased classification for face recognition," <http://arxiv.org/abs/1204.2358>.
- [35] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106–2112, 2010.
- [36] I. Naseem, R. Togneri, and M. Bennamoun, "Robust regression for face recognition," *Pattern Recognition*, vol. 45, no. 1, pp. 104–118, 2012.
- [37] M. A. Tahir, C. H. Chan, J. Kittler, and A. Bouridane, "Face recognition using multi-scale local phase quantisation and linear regression classifier," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 765–768, September 2011.
- [38] Y. Xu, A. Zhong, J. Yang, and D. Zhang, "LPP solution schemes for use with face recognition," *Pattern Recognition*, vol. 43, no. 12, pp. 4165–4176, 2010.
- [39] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.
- [40] Z.-Y. Liu, K.-C. Chiu, and L. Xu, "Improved system for object detection and star/galaxy classification via local subspace analysis," *Neural Networks*, vol. 16, no. 3-4, pp. 437–451, 2003.
- [41] J. Yang, L. Zhang, Y. Xu, and J.-Y. Yang, "Beyond sparsity: the role of L1-optimizer in pattern classification," *Pattern Recognition*, vol. 45, no. 3, pp. 1104–1118, 2012.
- [42] <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
- [43] [http://www.itl.nist.gov/iad/humanid/feret/feret\\_master.html](http://www.itl.nist.gov/iad/humanid/feret/feret_master.html)
- [44] <http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html>
- [45] X. F. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proceedings of the 19th Annual Conference on Neural Information Processing Systems*, pp. 507–514, MIT Press, 2005.

