Hindawi
Mathematical Problems in Engineering
Volume 2017, Article ID 1458412, 16 pages
https://doi.org/10.1155/2017/1458412

Research Article

Reconstructed Error and Linear Representation Coefficients Restricted by ℓ_1-Minimization for Face Recognition under Different Illumination and Occlusion

Xuegang Wu,[1,2] Bin Fang,[3] Yuan Yan Tang,[4] Xiaoping Zeng,[1] and Changyuan Xing[2]

[1]*College of Communication Engineering, Chongqing University, Chongqing 400030, China*
[2]*College of Computer Engineering, Yangtze Normal University, Fuling 408100, Chongqing, China*
[3]*College of Computer Science, Chongqing University, Chongqing 400030, China*
[4]*Faculty of Science and Technology, University of Macau, Macau*

Correspondence should be addressed to Xuegang Wu; xgwu@cqu.edu.cn

Received 25 November 2016; Accepted 23 April 2017; Published 22 May 2017

Academic Editor: Elisa Francomano

The problem of recognizing human faces from frontal views with varying illumination, occlusion, and disguise is a great challenge to pattern recognition. A general knowledge is that face patterns from an objective set sit on a linear subspace. On the proof of the knowledge, some methods use the linear combination to represent a sample in face recognition. In this paper, in order to get the more discriminant information of reconstruction error, we constrain both the linear combination coefficients and the reconstruction error by ℓ_1-minimization which is not apt to be disturbed by outliners. Then, through an equivalent transformation of the model, it is convenient to compute the parameters in a new underdetermined linear system. Next, we use an optimization method to get the approximate solution. As a result, the minimum reconstruction error has contained much valuable discriminating information. The gradient of this variable is measured to decide the final recognition. The experiments show that the recognition protocol based on the reconstruction error achieves high performance on available databases (Extended Yale B and AR Face database).

1. Introduction

Face recognition is an attractive field for researchers in computer vision and pattern recognition [1–4]. Different from other biometric techniques depending on cooperative subjects such as iris and fingerprint recognition, face images are usually obtained in unrestricted environments, which brings a serious challenge to face recognition methods. For example, as for the influence of illumination variations and occlusion, the appearances of a face complicate many problems in face recognition [5, 6]. Generally speaking, the decorations (such as sunglasses and scarves), and the shadows can be looked upon as different forms of occlusion which may lead to facial information losses [7]. Within the past decades, researchers made valuable contributions to the development of face recognition approaches under different illumination [8–12] and occlusion [13, 14], in which subspace learning as a class of important approaches is broadly adopted. Among those published articles related to subspace learning methods, two types of subspace learning methods, the global subspace learning (GSL) and the local subspace learning (LSL) (In the article, as we have known, GSL and LSL concepts are first proposed.), can be outlined. GSL method usually attains a subspace from all training samples, while LSL only forms one class of samples or a part of samples in different classes. For example, linear discriminant analysis (LDA) [15], marginal Fisher analysis (MFA) [16], principal component analysis (PCA) [17], maximum margin criterion (MMC) [18, 19], and independent component analysis (ICA) [20] fall in the category of GSL methods, and nearest subspace (NS) classification and common vector approach (CVA) [21] which only use class-specific samples as training samples

belong to LSL methods. LDA, MFA, and MMC are known to show better effects in ideal conditions. Otherwise, the performance of PCA and ICA is robust for the situation of pixel corruption to a certain degree. Nevertheless, when faces are occluded partially, the performance of these approaches will be degraded seriously, since the extracted features are caused to be unreliable.

Recently, the linear representations methods are applied in face recognition, which achieved high performance [22–25]. Sparse representation-based classification (SRC) [23], as one of the most representative approaches, will attain the best linear representation coefficients in a global sense, while another important method, linear regression-based classification (LRC) [22], tries to find linear representation coefficients in a local sense. The difference between SRC and LRC is that SRC uses all samples subspace, but LRC uses the sample subspace of each class to linearly represent a test sample. The subspaces in SRC and LRC are also different from above-mentioned subspace attained by subspace learning methods for only using original samples. If we look at linear representations method from a subspace perspective, linear representation method is also to find the best representation features in a global scope or a local one. In this sense, SRC can be treated as a GSL method and LRC as a LSL one. Usually, subspace learning methods will attain the subspace which benefits object recognition, such as the orthogonal subspace [15]. SRC only searches the sparse coefficients by using all training samples but not changing the original samples space, and it has demonstrated good potential in handling corruption and occlusion randomly in pixel and block. However, the method has been not robust to contiguous occlusion such as sunglasses and scarf [23]. Based on the working mechanism of SRC, Zhang et al. [25] recently proposed collaborative representation-based classification (CRC) approach. As well as SRC, LRC also does not change the original samples space. It is based on a model that samples from a specific object class lie on a linear subspace [15, 26] and uses class-specific training images to represent test image with only ℓ_2-norm constraint on fidelity term. The experiments on several databases have shown its efficacy and robustness to occlusion. In recent years, various learning methods are proposed, in which deep learning [27] shows its good performance in pattern recognition. However, other learning methods [28, 29] still devote their insight to this field.

In this paper, the constrained model based on the linear representations and class-specific concept is presented. From a subspace perspective, the method can be treated as a LSL one. And, in the model, both the linear combination coefficients and the minimum reconstruction error are restricted by ℓ_1-minimization. Usually, a reconstruction error which is generated from the true identity will reflect only intraclass difference. When it is attained by nontrue subjects, the reconstruction error will contain not only the intraclass difference but also much detailed information from the interclass difference. However, in a complicated environment, the reconstruction error may be affected by the pixel corruption caused by random noise, face expression, illumination variations, and occlusion. In our paper, the constrained reconstruction error can keep the similarity in the intraclass and the difference in the interclass as much as possible under complicated environment. And, we believe that the reconstruction error in the model includes much discriminating information which is of benefit to face recognition.

Our main advantages are listed as follows:

(1) We propose a constrained linear model which is of benefit to face recognition under different degree of lighting and occlusion.

(2) The objective function is transformed into a form which can be tackled by a convex optimization method.

(3) In the approach, the assumption of explicit prior knowledge about the sources of light is not assigned, as well as the feature of regions corrupted and occluded.

(4) Not only is our method used in face pattern recognition, but it can also be applied to other image-based object identification.

The rest of the paper is organized as follows. In Section 2, we show the analysis and motivation of our approach. Then, the proposed algorithm is shown in Section 3. In Section 4, further experiments show the robustness to occlusion and illumination in face recognition by comparison with some classic recognition techniques. Finally, Section 5 concludes the study.

2. Analysis and Motivation

In the study of face recognition, illumination variations, occlusion, and disguise have been challenging difficulties compromising the recognition performance. It has been proven that, for raw images, variations caused by illumination and partial occlusion are more significant than the inherent differences between individuals [30]. Shadows and partial occlusion can be considered as spatial errors in face images. In recent decades, there are numerous applications where the data under study can be represented by the linear representation methods from the class-specified samples or all classes ones, which naturally fit a linear representation model. So, when we have a number of face images of an individual, it is reasonable to believe that the linear representation methods have the advantage in removing these errors.

Figure 1 shows example images from the database of Extended Yale B [31, 32]. The resolution of every image is 30×28. We randomly choose 8 samples under varying illumination but no occlusion, which we stack as the columns of matrix $D \in \mathbb{R}^{840\times 8}$. From Figure 2, It can be seen clearly that although LRC used class-specific samples to linearly represent probe face image, under the constraint of least square method, the linear representation coefficients may be overfit in the conditions of different illumination and occlusion [23]. Thus, the reconstruction error will be led to be far from the true value. Figures 2(a) and 2(b) show separately the difference of the reconstruction error based on same class samples (Figure 2(a), top level images) and other class (Figure 2(b), bottom level images) in LRC.

Figure 1: Five subgroups in Extended Yale B database. The top-down view, every row expresses samples from subgroups, separately.

All coefficients of training samples are constrained sparsely to represent a probe sample in SRC. However, many linear representation coefficients from other classes samples will also unavoidably emerge. Figures 3(a) and 3(b) is the test face image and training faces in the AR Face database which will be introduced later. In Figure 4(b), left and mid images show that the reconstruction images are separately attained by the product of the class-specific samples with their linear representation coefficients and the product of the rest samples with their coefficients. In Figure 4(b), right image shows the difference between two reconstruction images. When the reconstruction error in a global sense is formed by all training samples, it is almost not of benefit to face recognition in the conditions of occlusion and illumination. So, the recognition protocol in SRC only uses the best linear representation coefficients.

In Figure 2, although the reconstruction error in LRC shows much difference when using the training samples to linearly represent a probe sample from same class and another class, the linear representation coefficients when representing different probe samples are not salient. A new insight into face recognition is that a reconstruction error is vital to face recognition under different occlusion and illumination. It inspired us to attain the accurate reconstruction error by designing a new method. The model is shown in formula (1). In Figure 4(c), it can be seen that the reconstruction error is obvious, in which the left image is from same class and the right one from another class. We can see that the texture of the right one is more obvious than the left, which expresses the gradient value with obvious texture being bigger than one with nonobvious texture. In Figure 5, candidates were sampled randomly from the 15th individual of Subset 1 for 50 times in Yale B and the average gradient values of the reconstruction error is highlighted in red. In the figure, the lowest gradient value among the 38 classes is attained by the proposed algorithm. In the experiments, the results show that if the test sample and the training samples are with the same label, the reconstruction error gradient value attained is much lower than the value when the test sample and the training samples are with the different label. And our method achieves 100 percent recognition rate in the first subset of the Extended Yale B database. In Figure 6, the reconstruction error and the linear representation coefficients are attained through solving the proposed model. From Figures 6(a), 6(b),

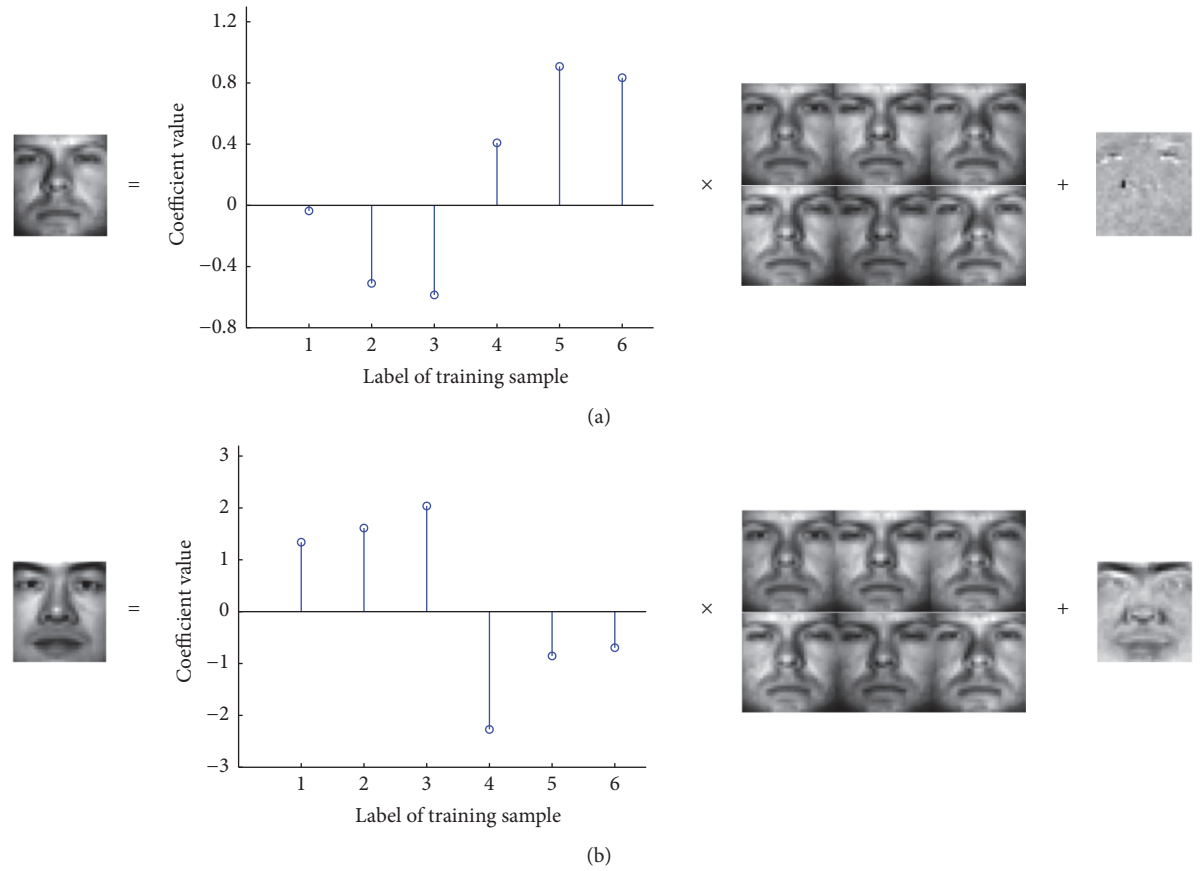


Figure 2: LRC experiments on Extended Yale B database. (a) Representing the probe image with the samples from the same class in LRC. (b) Representing the probe image from other class with the samples from the above same class in LRC.

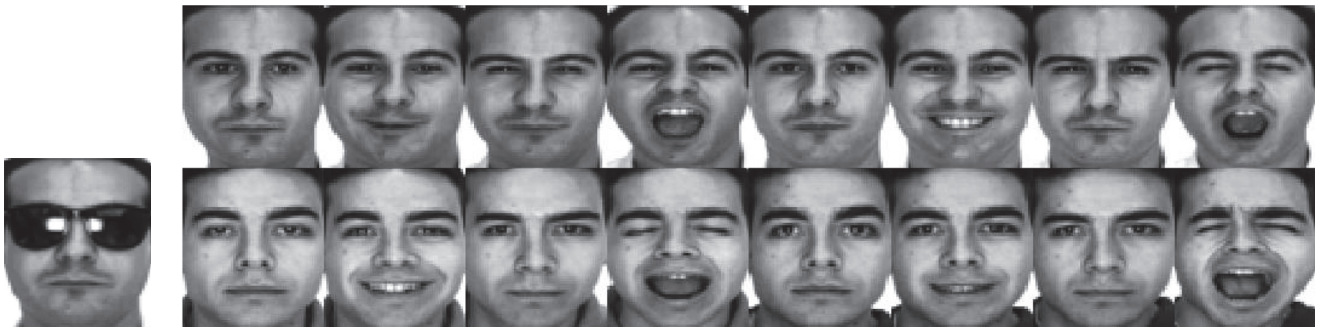
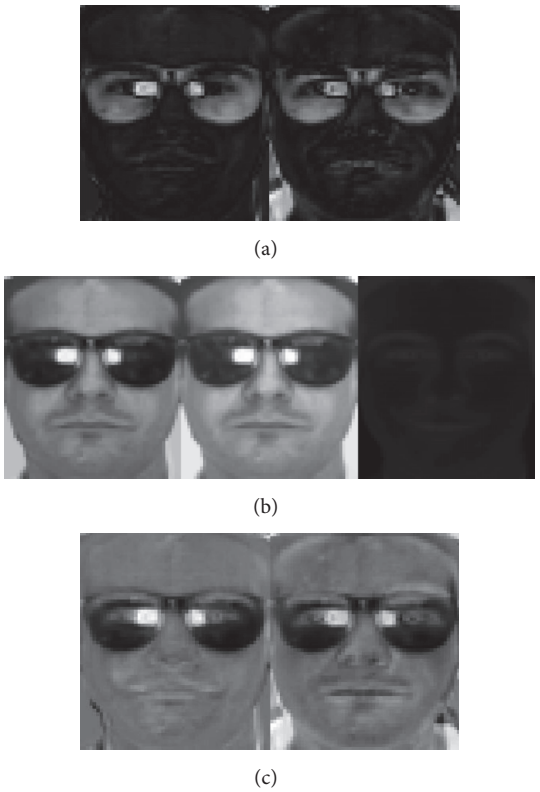


Figure 3: Example faces. (a) Test image. (b) Training images.

7(a), and 7(b), we can see that it is different between these two linear representation coefficients figures, in which one represents the probe image with the samples from the same class and the other represents the probe image with another class. In particular, from Figures 8 and 9, it is can be clearly seen that the variation scope of representation coefficients value in LRC is much smaller than the proposed approach. In Figure 9, we also observe that these two variation scopes of representation coefficients value in the proposed method are very big, in which one is from the same class representation and the other from other class.

Next, why can we attain the relatively accurate reconstruction error? The analysis is as follows.

(A) Linear Representation-Based Method. We first suppose the linear representation coefficients to be $x \in \mathbb{R}^{m\times 1}$, the training samples' matrix $A \in \mathbb{R}^{n\times m}(n \gg m)$ (in which, m denotes the number of training samples and n the dimension of each



(a)

(b)

(c)

Figure 4: The comparison of three different algorithms based on Figure 2. (a) Left: reconstruction error image with its own class in LRC. Right: reconstruction error image with other class in LRC. (b) Left: reconstruction error image with its own class in SRC. Middle: reconstruction error image in other samples in SRC. Right: residual image between left image and mid in SRC. (c) Left: the reconstruction error image of the proposed method with its own class. Right: the reconstruction error image of the proposed method with the other class.

training sample), and the linear representation error $\varepsilon \in \mathbb{R}^{n\times 1}$ and $y \in \mathbb{R}^{n\times 1}$ is the reconstruction sample. The general function can be represented as follows:

$$y = Ax + \varepsilon. \tag{1}$$

(B) SRC Method. The reconstruction sample of y is composed of two parts in SRC, one is the component $A_i x_i$ of the ith class; another is $A_{\neq i}x_{\neq i}$ from other classes. x_i and $x_{\neq i}$ are the sparse representation coefficients separately from the ith class and other classes. ε_g denotes the global reconstruction error vector of SRC. So, y can be linearly represented as follows:

$$y = A_i x_i + A_{\neq i}x_{\neq i} + \varepsilon_g. \tag{2}$$

In SRC, the reconstruction error is

$$\varepsilon_g = y - A_i x_i - A_{\neq i}x_{\neq i}, \quad A_{\neq i}x_{\neq i}(:) \ll A_i x_i(*)$$
$$\left(\sum_j x_j > 0, \ \ j = \{1, 2, \ldots, i-1, i+1, \ldots, n\}\right), \tag{3}$$

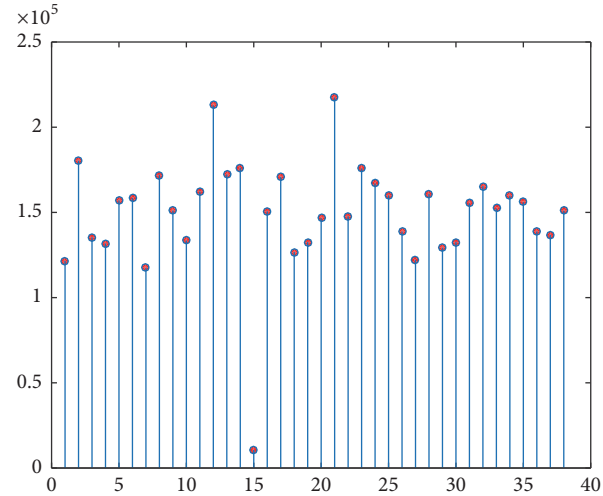


Figure 5: A validation of the proposed method, candidates sampled randomly from the 15th individual of Subset 1 for 50 times in Yale B and the average gradient of reconstructed error image is highlighted in red.

in which $(*)$ denotes all elements in a vector. So, the reconstruction error component ε_g is not only from the same class but from other classes.

(C) LRC Method. In LRC, the linear representation formula is

$$y = A_i x_i + \varepsilon_i, \tag{4}$$

in which A_i is the gallery samples from the ith class, x_i is the linear representation coefficients on A_i, and y is the reconstruction sample.

For the least squares method that may lead to overfitting,

$$\min \quad \|y - A_i x_i\|_2$$
$$\varepsilon_i = y - A_i x_i \pm ov_i, \quad |ov_i(:)| \ll |A_i x_i(*)|, \tag{5}$$

in which $(*)$ is the same with the above, ov_i represents as an overfitting variable from the ith class, and $\|\cdot\|_2$ denotes the ℓ_2-norm.

(D) The Proposed Method. From (B) and (C), the above two methods will bring more or less information loss. In image-based recognition, ℓ_1-norm can be robustly represented as a probe sample from class-specific samples and is not easy to be affected by the outliers. Thus, the training samples that is class-specific (from its same class) will represent the test sample well. In Figure 6, the linear representation coefficients show that the values are with the sparse feature. The phenomenon is caused by the constraints of ℓ_1-norm which forces the probe sample to be represented by the most similar ones. And, through observing error variable ε_i restricted by ℓ_1-norm, we know that the reconstruction error also shows sparse feature when the probe sample is classified into its own class.

From above formulas (3) and (5), the error components ε_g and ε_i in SRC and LRC separately are not satisfied

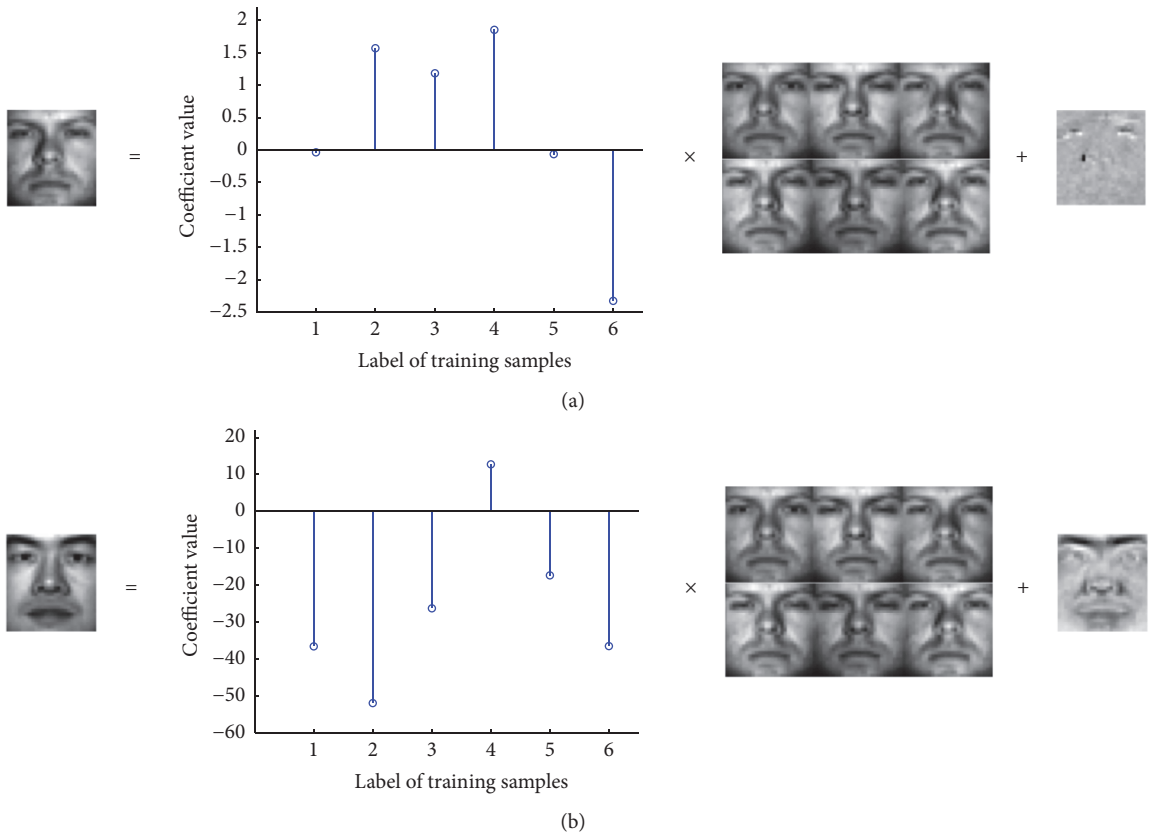


Figure 6: Proposed approach on Extended Yale database. (a) Representing the probe image with the samples from the same class by the proposed approach. (b) Representing the probe image from another class with the samples from other class (not same class with the probe image) by the proposed approach.

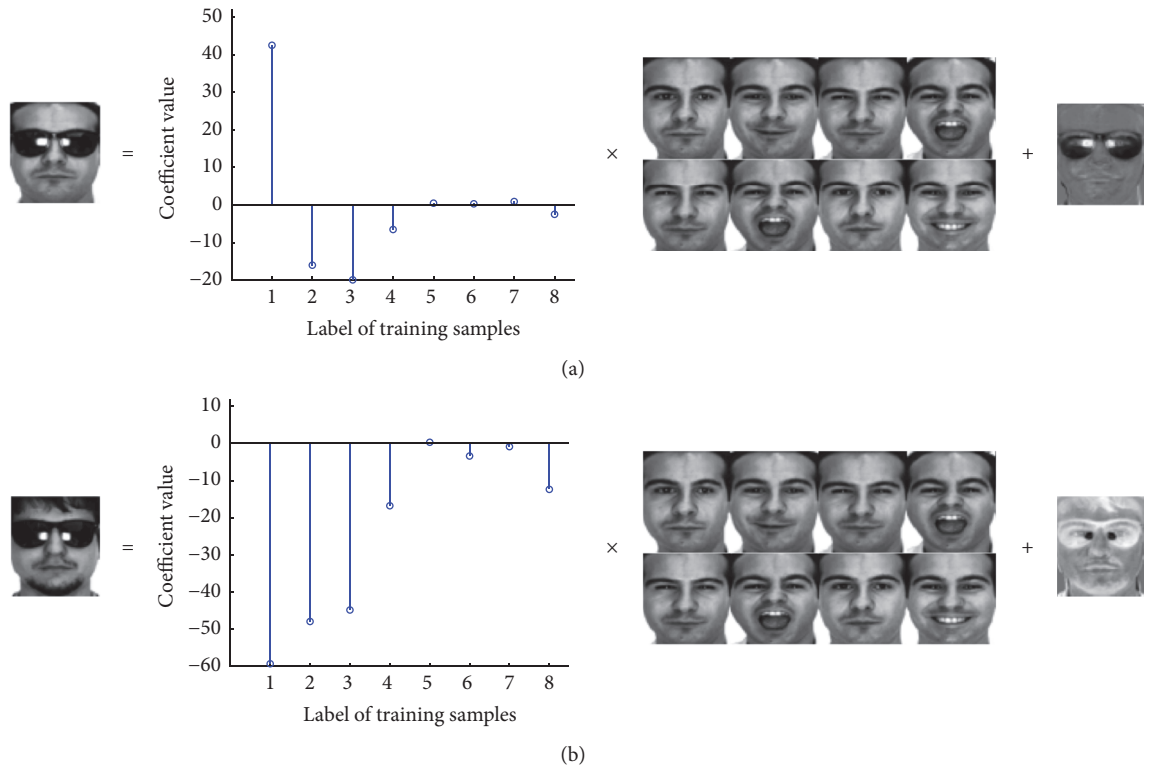


Figure 7: Proposed approach on AR database. (a) Representing the probe image with the samples from the same class by the proposed approach. (b) Representing the probe image from other class with the samples from other class by the proposed approach.

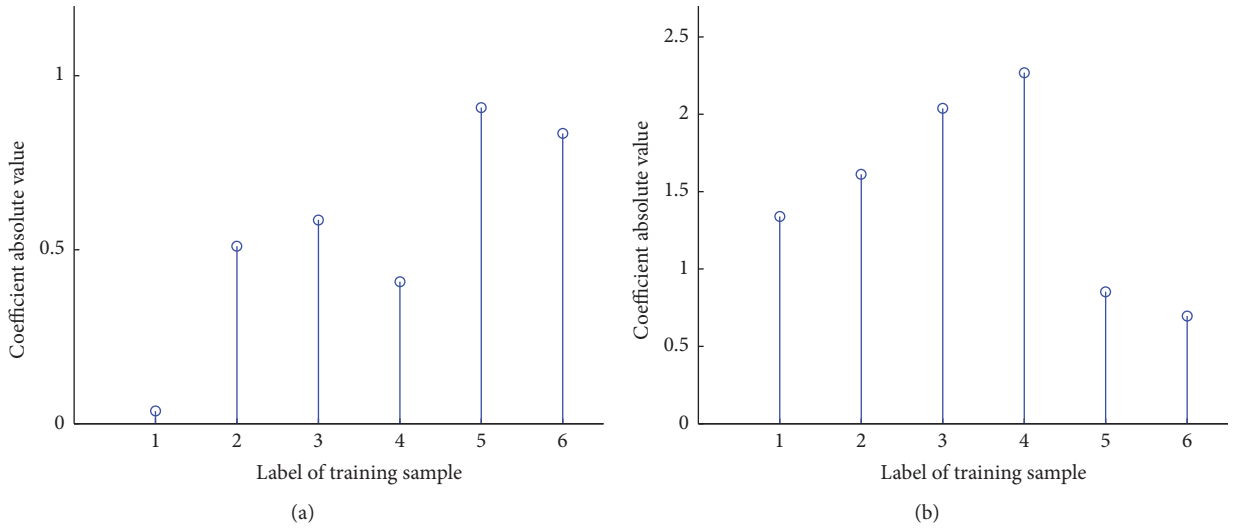


Figure 8: The absolute linear representation coefficients are (a) represented by the samples of the same class and (b) represented by the samples of another class in LRC.

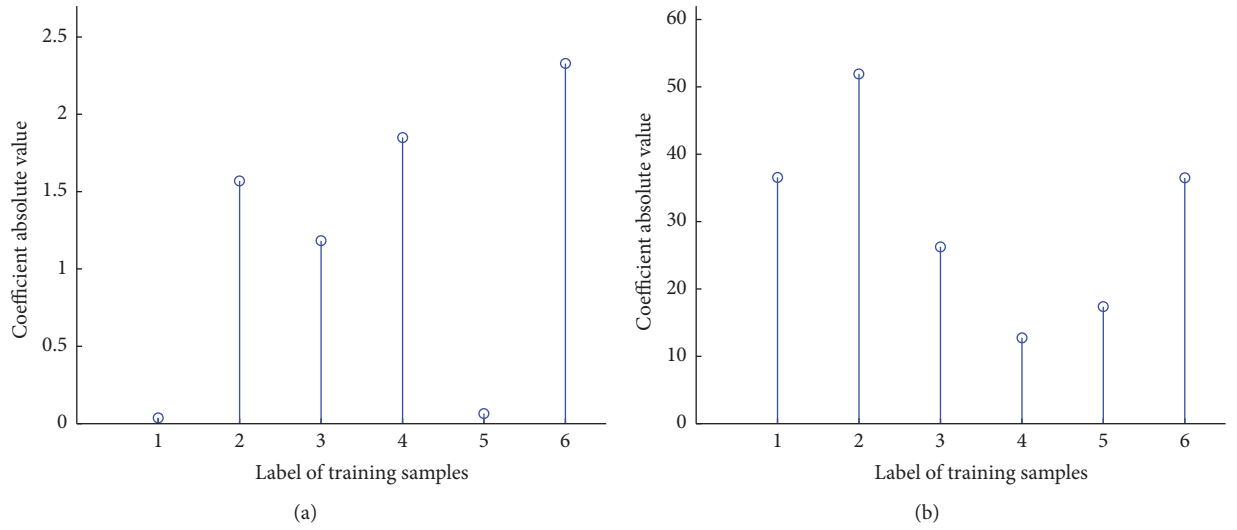


Figure 9: The absolute linear representation coefficients are (a) represented by the samples of the same class and (b) represented by the samples of another class in the proposed method.

for further recognition. We argue that more accurate error component has much more discriminative power than the above approaches in the image-based recognition. Next, how to represent the error component as accurately as possible becomes the main problem placed in front of us.

In the proposed method, the reconstruction error ε_i is represented by its own training samples as accurately as possible, and we also enforce the ℓ_1-norm on it. To represent the error component as accurately as possible, the objective function is as follows:

$$\begin{aligned} \min \quad & \|x_i\|_1 \,\&\& \\ \min \quad & \|\varepsilon_i\|_1, \\ \text{s.t.} \quad & y = A_i x_i + \varepsilon_i, \end{aligned} \tag{6}$$

in which && is the logical intersection, x_i is the linear representation coefficients from the ith class, y is the reconstruction sample, and $\|\cdot\|_1$ denotes the ℓ_1-norm. If the vector ε_i is attained when the probe face image is not from its same class, the elements in the vector will be not sparse and when the vector is transformed into a matrix, its gradient value will become bigger. And, when some outliers (the variation of a sample is big, such as different expression or illumination) mix into the same class, under the constraints of ℓ_1-norm, the sparse representation coefficients on the outliers will also approach zeros, and the probe sample will still be represented as accurately as possible in our method. From Figure 6, we can see that the coefficients become sparse, it implies that the probe image will be linearly represented with the higher

Input: The training samples $S = [S_1 \ S_2 \ \cdots \ S_c] \in \mathbb{R}^{m\times n}$ for c classes, $S_i \in \mathbb{R}^{m\times n_i}$ denotes the training sample matrix from ith class, n_i denotes the number of training images of the ith individuals. A test sample $y \in \mathbb{R}^{m\times 1}$.

(1) $\min \|w_1\|_1$ && $\min \|w_2\|_1$, s.t. $y = S_i w_1 + w_2$.

(2) $y = S_i w_1 + w_2 = S_i w_1 + I w_2 = [S_i, I] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$,

(3) Let $B_i = [S_i \ \ I]$, $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$.

(4) Compute: $\widehat{w} = \arg\min \|w\|_1$, s.t. $y = B_i w$.

(5) **for** each subject i **do**

(a) Use formula (13)–(29) to solve the objective function

$$\arg\min \|w\|_1 \ \text{ s.t. } y = B_i w,$$
$$B_i = [S_i, I], \ I = \text{eye}(m),$$

(b) the solution is

$$\widehat{w} = \begin{bmatrix} \widehat{w}_1 \\ \widehat{w}_2 \end{bmatrix},$$

(c) the unknown variables are

$$w_1 \longleftarrow \widehat{w}_1$$
$$w_2 \longleftarrow \widehat{w}_2.$$

(6) **end for**

(7) Calculate the ℓ_2-norm of w_1,

(8) Calculate the Gradient of $\text{img}(w_2)$.

Output: the recognition protocols:

Reconstruction error based method,

$$\text{indentify}(R_r(E)) = \min_i(\text{Gra}(\text{img}(w_2)_i)).$$

Algorithm 1: The proposed method.

similar degree samples. The above analysis clearly illustrates that our method has the stable performance.

3. Proposed Method

3.1. Algorithmic Description. Assume n training samples for all c individuals in a face dataset, and m samples are available for each individual (actually, the number of training images for each identity might be different) with a resolution of m = width × height pixels. n_i indicates the number of training images of the ith individuals. In this paper, each image is converted into a long vector by stacking its columns one by one; S_i is a matrix with $m \times n_i$ dimension (Algorithm 1). Then, the images in the c subjects are arranged in the following matrix:

$$S = [S_1, S_2, \ldots, S_c] \in \mathbb{R}^{m\times n}, \tag{7}$$

in which S is a matrix consisting of the training sample vectors. A test sample is represented by $y \in \mathbb{R}^{m\times 1}$. Considering the following, optimization problem can be described as follows:

$$\begin{aligned} \min \quad & \|w_1\|_1 \ \&\& \\ \min \quad & \|w_2\|_1, \\ \text{s.t.} \quad & y = S_i w_1 + w_2, \end{aligned} \tag{8}$$

in which && is the logical intersection. In order to get the optimal solution, we transform the objective function into the following:

$$y = S_i w_1 + w_2 = S_i w_1 + I w_2 = [S_i, I] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \tag{9}$$

in which $I \in \mathbb{R}^{m\times m}$ represents identity matrix. Then, the problem becomes

$$\begin{aligned} \widehat{w} = \arg\min \quad & \|w\|_1, \\ \text{s.t.} \quad & y = B_i w, \end{aligned} \tag{10}$$

in which $B_i = [S_i \ \ I] \in \mathbb{R}^{m\times(n_i+m)}$, and $m < n_i + m$, $S_i \in \mathbb{R}^{m\times n_i}$ denotes the training sample matrix of ith class, and $I \in \mathbb{R}^{m\times m}$ is an identity matrix which used to express potential occlusion. A part of $\widehat{w}$, entries from $n_i + 1$ to $n_i + m$, represents linearly occlusion coefficients. And $w = \left[\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right] \in \mathbb{R}^{(n_i+m)\times 1}$, $w_1 \in \mathbb{R}^{n_i\times 1}$, $w_2 \in \mathbb{R}^{m\times 1}$. The training set is composed of complete nonoccluded face images. Figures 5(a) and 6(a) show that a partially occluded face can be sparsely represented by the training samples from class-specified samples, and face data actually is distributed in low-dimensional subspace and satisfies the linear equation constraints. The difference of the linear representation coefficients in Figure 7 further verifies the above claims. Whether a probe image may be partially occluded or not, the error caused by occluded pixels can only be squeezed into the identity matrix $I \in \mathbb{R}^{K\times K}$. The pixels,

corresponding to the entries of $\widehat{w}(n_i + 1 : n_i + m)$ with big values, have extremely high probability to be occluded.

Nowadays, improvement in the theory of sparse representation and compressed sensing [33–35] reveals that if the solution $\|w_1\|_0$ to achieve is fairly sparse, the solutions of the ℓ_0-minimization problem (11) and the ℓ_1-minimization problem are equivalent. On the contrary, in some degree, we also believe that the ℓ_1-minimization solution is with sparse feature.

3.2. The Solution of the Proposed Algorithm. We will solve the following objective function to achieve the optimal results;

$$\begin{aligned} \arg\min \quad & \|w\|_1 \\ \text{s.t.} \quad & y = B_i w. \end{aligned} \tag{11}$$

When w, B_i, and y are real number, (11) is looked upon as a linear program question [36]:

$$\begin{aligned} \min_{w,u} \quad & \sum_i u_i \\ \text{s.t.} \quad & w_i - u_i \le 0, \\ & -w_i - u_i \le 0, \\ & B_i w = y, \end{aligned} \tag{12}$$

in which $A \in \mathbb{R}^{p\times n}$ and its rank is $p < n$. This solution of above objective function is attained by a standard primal-dual method.

Set

$$\begin{aligned} f_{u_\alpha;i} &\coloneqq w_i - u_i, \\ f_{u_\beta;i} &\coloneqq -w_i - u_i. \end{aligned} \tag{13}$$

We define the *Lagrangian* L associated with the above objective function:

$$\begin{aligned} L\left(w, u, v, \lambda_{u_\alpha}, \lambda_{u_\beta}\right) = & \sum_i u_i + \sum_i \lambda_{u_\alpha;i} f_{u_\alpha;i} \\ & + \sum_i \lambda_{u_\beta;i} f_{u_\beta;i} + v^T \left(B_i w - y\right). \end{aligned} \tag{14}$$

Suppose $f_{u_\alpha} = \left(f_{u_\alpha;1} \ \cdots \ f_{u_\alpha;N}\right)^T$ and $f_{u_\beta}, \lambda_{u_\alpha}, \lambda_{u_\beta}$ are like that form. And

$$\begin{aligned} \nabla f_{u_\alpha;i} &= \begin{pmatrix} \delta_i \\ -\delta_i \end{pmatrix}, \\ \nabla f_{u_\beta;i} &= \begin{pmatrix} -\delta_i \\ -\delta_i \end{pmatrix}, \\ \nabla^2 f_{u_\alpha;i} &= 0, \\ \nabla^2 f_{u_\beta;i} &= 0, \end{aligned} \tag{15}$$

in which δ_i denotes the standard basis vector of component i. Let $w^\star$, $u^\star$, $\lambda^\star_{u_\alpha}$, $\lambda^\star_{u_\beta}$, and $v^\star$ be any points which are zero duality gap, and these points are primal and dual optimal points. As $w^\star$ minimizes the objective function $L(w, u^\star, v^\star, \lambda^\star_{u_\alpha}, \lambda^\star_{u_\beta})$ on x, $u^\star$ minimizes it on u. Then, it shows that the gradient must disappear at $w^\star$, $u^\star$; that is,

$$\begin{aligned} & \left.\frac{\partial L\left(w, u, v, \lambda_{u_\alpha}, \lambda_{u_\beta}\right)}{\partial w}\right|_{u=u^\star, w=w^\star, \lambda_{u_\alpha}=\lambda^\star_{u_\alpha}, \lambda_{u_\alpha}=\lambda^\star_{u_\beta}, v=v^\star} \\ & = \lambda^\star_{u_\alpha} - \lambda^\star_{u_\beta} + A^T v^\star = \mathbf{0}, \\ & \left.\frac{\partial L\left(w, u, v, \lambda_{u_\alpha}, \lambda_{u_\beta}\right)}{\partial u}\right|_{u=u^\star, w=w^\star, \lambda_{u_\alpha}=\lambda^\star_{u_\alpha}, \lambda_{u_\alpha}=\lambda^\star_{u_\beta}, v=v^\star} \\ & = \mathbf{1} - \lambda^\star_{u_\alpha} - \lambda^\star_{u_\beta} = \mathbf{0}. \end{aligned} \tag{16}$$

The KKT conditions are listed as follows:

$$\begin{aligned} \lambda^\star_{u_\alpha} - \lambda^\star_{u_\beta} + A^T v^\star &= \mathbf{0}, \\ \mathbf{1} - \lambda^\star_{u_\alpha} - \lambda^\star_{u_\beta} &= \mathbf{0}, \\ B_i w^\star - y &= \mathbf{0}, \\ \lambda^\star_{u_\alpha;i} f_{u_\alpha;i} &= 0, \\ \lambda^\star_{u_\beta;i} f_{u_\beta;i} &= 0, \\ f_{u_\alpha;i}, f_{u_\beta;i} &\le 0, \\ \lambda^\star_{u_\alpha;i}, \lambda^\star_{u_\beta;i} &\ge 0, \end{aligned} \tag{17}$$

where $i = 1, \ldots, n$. Thus, we get the central and dual residuals:

$$\begin{aligned} \text{res}_{\text{cent}} &= \begin{pmatrix} -\Lambda_{u_\alpha} f_{u_\alpha} \\ -\Lambda_{u_\beta} f_{u_\beta} \end{pmatrix} - \left(\frac{1}{\tau}\right)\mathbf{1}, \\ \text{res}_{\text{dual}} &= \begin{pmatrix} \lambda_{u_\alpha} - \lambda_{u_\beta} + A^T v \\ \mathbf{1} - \lambda_{u_\alpha} - \lambda_{u_\beta} \end{pmatrix}. \end{aligned} \tag{18}$$

Then, the original *complementary slackness* condition is $\lambda_i f_i = 0$. Now, it is relaxed to

$$\lambda^k_i f_i\left(z^k\right) = -\frac{1}{\tau^k}, \tag{19}$$

where the parameter τ^k is judiciously increased when progressing through the Newton iterations; diagonal matrix Λ is set to $(\Lambda)_{ii} = \lambda_i$. KKT conditions can be denoted as $r_t(w, u, v) = 0$, in which it can be expressed as

$$r_t(w, u, v) = \begin{pmatrix} \lambda_{u_\alpha} - \lambda_{u_\beta} + A^T v \\ \mathbf{1} - \lambda_{u_\alpha} - \lambda_{u_\beta} \\ B_i w - y \end{pmatrix}. \tag{20}$$

Use the following Newton step Δz to solve $r_t(x, u, v) = 0$ at the current point:

$$z = (w, u, v), \quad (\Delta z = (\Delta w, \Delta u, \Delta v)). \tag{21}$$

The following linear equation expresses Newton step:

$$r_t(z+\Delta z) \approx r_t(z) + Dr_t(z)\Delta z = 0. \tag{22}$$

Then, $\Delta z = -Dr_t(z)^{-1} r_t(z)$. In terms of w, u, and v, we have

$$\begin{pmatrix} \Sigma_1 & \Sigma_2 & A^T \\ \Sigma_2 & \Sigma_1 & 0 \\ A & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta w \\ \Delta u \\ \Delta v \end{pmatrix} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{pmatrix} \coloneqq \begin{pmatrix} \left(-\frac{1}{\tau}\right) \cdot \left(-f_{u_\alpha}^{-1} + f_{u_\beta}^{-1}\right) - A^T v \\ -\mathbf{1} - \left(\frac{1}{\tau}\right) \cdot \left(f_{u_\alpha}^{-1} + f_{u_\beta}^{-1}\right) \\ y - B_i w \end{pmatrix}. \tag{23}$$

Let the primal-dual search direction $\Delta y = (\Delta w_{\text{pds}}, \Delta u_{\text{pds}}, \Delta v_{\text{pds}})$ be the above formula solution. In the above formula,

$$\begin{aligned} \Sigma_1 &= \Lambda_{u_\alpha} F_{u_\alpha}^{-1} - \Lambda_{u_\beta} F_{u_\beta}^{-1}, \\ \Sigma_2 &= \Lambda_{u_\alpha} F_{u_\alpha}^{-1} + \Lambda_{u_\beta} F_{u_\beta}^{-1}, \end{aligned} \tag{24}$$

in which $F_\bullet$ denotes diagonal matrices, $(F_\bullet)_{ii} = f_{\bullet;i}$, and $f_{\bullet;i}^{-1} = 1/f_{\bullet;i}$.

Set

$$\Sigma_w = \Sigma_1 - \Sigma_2^2 \Sigma_1^{-1}, \tag{25}$$

The following items will be removed:

$$\begin{aligned} \Delta w &= \Sigma_w^{-1}\left(\varphi_1 - \Sigma_2\Sigma_1^{-1}\varphi_2 - A^T\Delta v\right) \\ \Delta u &= \Sigma_1^{-1}\left(\varphi_2 - \Sigma_2 \Delta w\right). \end{aligned} \tag{26}$$

and solve

$$-A\Sigma_w^{-1}A^T\Delta v = \varphi_3 - A\left(\Sigma_w^{-1}\varphi_1 - \Sigma_w^{-1}\Sigma_2\Sigma_1^{-1}\varphi_2\right). \tag{27}$$

Equation (28) is $p \times p$ equations. It is also a positive definite system. We can solve it by conjugate gradients.

Δw, Δu, Δv are attained; then compute the change:

$$\begin{aligned} \Delta\lambda_{u_\alpha} &= \Lambda_{u_\alpha} F_{u_\alpha}^{-1}(-\Delta w + \Delta u) - \lambda_{u_\alpha} - \left(\frac{1}{\tau}\right) f_{u_\alpha}^{-1} \\ \Delta\lambda_{u_\beta} &= \Lambda_{u_\beta} F_{u_\beta}^{-1}(\Delta w + \Delta u) - \lambda_{u_\beta} - \left(\frac{1}{\tau}\right) f_{u_\beta}^{-1}. \end{aligned} \tag{28}$$

3.3. Analysis of Classification. The reconstruction error is the component which cannot be represented linearly. It may come not only from the normal noise but also from occlusion or illumination. If the error can be attained accurately, it will play an important role in face recognition because the test sample with a lower gradient value is inclined to be the same label of the training samples. We will first look at w_1 and w_2 as follows.

In the proposed objective function, w_1 and w_2 represent the linear representation coefficients and the reconstruction error separately. In the paper, the constraints can make them become sparse through the optimization solution. That is, w_1 is the linear representation coefficients with the sparse character and w_2 is the sparse reconstruction error. When the label of the training set and the test sample is the same, the ℓ_1-norm value of w_1 will be smaller than the value when the label is different, which is shown in Figures 6 and 7.

In the paper, the gradient of $\text{img}(w_2)$ is regarded as the recognition protocol. The probe image represented linearly by its true class is similar to its own gallery samples; then the gradient of $\text{img}(w_2)$ will be very small. Thus, the error component including the noise and occlusion is not very obvious. The major variance may come from the occlusion, noise, and expression. Thus, we note that the correlation error between the gallery samples and the probe sample is vital to the final classification.

3.4. Classification Rule. These above considerations lead us to summarize the overall formula as follows.

First, in vector $w = \left[\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right]$, we take w_2 out and reshape the vector to be an image matrix as the probe image. And, the reconstruction error image will be denoted by $\text{img}(w_2)$.

Here, the recognition protocols based on the gradient of reconstruction error are defined as the following forms.

Reconstruction error based method is

$$\text{indentify}\left(R_r(E)\right) = \min_i \left(\text{Gra}\left(\text{img}\left(w_2\right)_i\right)\right), \tag{29}$$

in which $\text{Gra}(*)$ denotes the gradient of $*$.

4. Experimental Results

In this section, the experiments are presented on publicly available databases with different illumination, occlusion, and disguise. And the experiments are designed to show the efficacy of the our approach, which are implemented by Matlab R2011b on a desktop running Windows XP with Intel Pentium Dual-Core processor 2.60 GHz CPU. The Extended Yale B database [31, 32] and AR database [37] are usually used as the benchmarks to validate various algorithms. So, our paper also uses both of them. First, we will compare the proposed method with several classical methods on addressing different illuminations. Then, the experiments will show the robustness of our method for random pixel corruption and random block occlusion, respectively. Finally, we observe the performance in our method when the samples in disguise are from the AR Face database, respectively. In addition, since feature selection may affect classification results, our method directly works on original images without feature extraction. We choose the ℓ_1-norm method in the paper to solve the ℓ_1-regularized minimization in SRC. In the paper, we will show the efficiency of reconstruction error based method.

4.1. Recognition under Varying Illumination. The Extended Yale B database consists of 2,414 frontal face images of 38 individuals under various laboratory controlled lighting conditions. All test image data used in the experiments are manually aligned, cropped, and then resized to 30×28

Table 1: Correct recognition (%) for the database of Extended Yale B with subset 1 as training set.

Subsets	2	3	4	5
Eigenfaces + NN	89.04	30.29	5.48	3.08
Fisherfaces + NN	100	85.52	21.49	4.06
Gradientfaces	100	93.90	60.75	55.32
SRC	100	98.86	67.32	22.27
LRC	100	98.67	86.84	37.25
CRC-RLS	99.78	99.05	88.16	40.48
Our method$_{error}$	100	99.62	89.69	48.04

images. The database is composed of five subsets (see Figure 1): Subset 1 consisting of 266 images under normal illumination conditions and each subject including seven images; Subset 2 and Subset 3 under slight-to-mode rate illumination variations, and each subset including 12 images, while Subset 4 including 14 images in a subject and Subset 5 including 19 images under severe light variations. In this experiment, Subset 1 is chosen for training, and other subsets are chosen for testing.

Table 1 shows the recognition performance for varying subsets. Eigenfaces and Fisherfaces are combined with Nearest Neighbors (NN) classifier for classification. We also compare the proposed method with CRC of regularized least square (CRC-RLS) [25] and Gradientfaces [9], in light of its capability of handling illumination changes. Note that Fisherfaces + NN, LRC, SRC, Gradientfaces, and the proposed methods show excellent performance for moderate light variations, yielding 100% recognition accuracy for Subsets 2. CRC-RLS also achieves 99.78%. In the case of severe light variations, recognition accuracy of SRC and LRC, however, falls to 67.32% and 86.84% for Subsets 4, respectively. Our method achieves 89.69% on Subset 4, outperforming the best competitor CRC-RLS by a margin of 1.53%. For Gradientfaces, it is derived from the image gradient domain such that it can discover underlying inherent structure of face images since the gradient domain explicitly considers the relationships between neighboring pixel points [9]. We can see that Gradientfaces achieves stable recognition accuracy for all subsets, which demonstrates its ability of extracting illumination insensitive features. In particular, Gradientfaces obtains the best recognition rate of 55.32% on Subset 5. By contrast, recognition accuracy of the other approaches drops heavily in this subset, while the proposed method still outperforms LRC, SRC, CRC-RLS, and the benchmark approaches of Eigenfaces + NN and Fisherfaces + NN.

Notice that all rival methods attained poor performance on the most challenging subset, that is, Subset 5. The proposed method achieves 99.62% recognition rate on the second most challenging subset, and 89.69% drops to 48.04% for the most challenging subset. For Subset 5, cast shadows and the presence of severe illuminations (e.g., overhead and lateral and even rear illumination) are some of the difficulties of note. The characteristic facial features, such as the eyes and the nose, are typically of great importance for recognition. To substantiate the explanation of the performance of our method, Subset 1 and Subset 2 are chosen for training and testing, respectively. Three face features (nose, eyes, and mouth) will be occluded by using a black rectangle to simulate the occlusion. See Figure 10 for an example of the occluded face parts. The results in Table 2 show that our method reaches 100% accuracy for all cases. The results show that the comparable algorithms except Eigenfaces cope well with block occlusion.

Table 2: Correct recognition (%) on Subset 2 with partial occlusion in the Extended Yale B database with Subset 1 as training set.

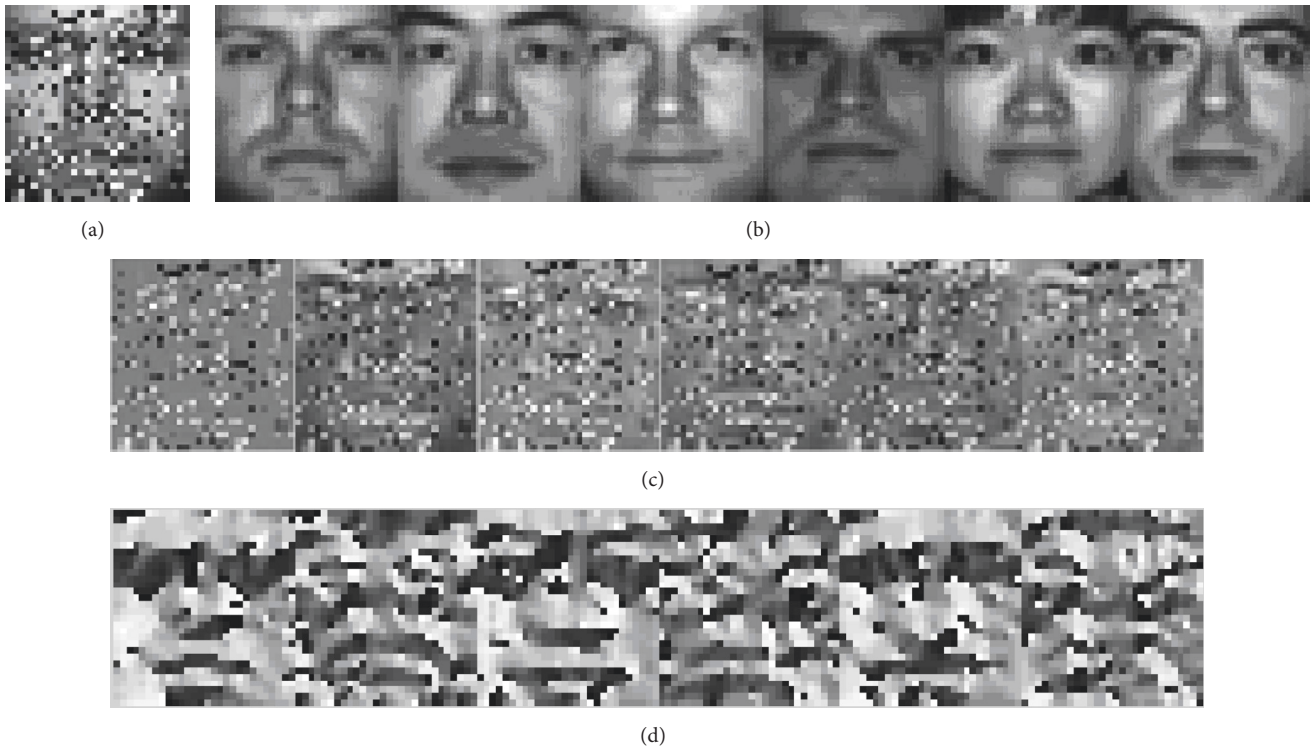
Occluded region	Mouth	Nose	Eyes
Eigenfaces + NN	87.94	89.69	86.62
Fisherfaces + NN	100	100	100
SRC	100	100	100
LRC	100	100	100
Our method$_{error}$	100	100	100

4.2. Recognition with Random Pixel Corruption. In this experiment, we test the robustness of our method in the presence of pixel corruption. We use Subset 1 for training and Subsets 2 and 3 for testing, respectively. For each testing image, we replace a certain percentage of its pixels with uniformly distributed random values within $[0, 255]$. The corrupted pixels are randomly chosen for each test image and their locations are unknown to the proposed algorithm. For Subset 2, we adjust the percentage of corrupted pixels from 10 percent to 90 percent to check the efficiency of the algorithm. Figure 11(a) is an example of a test image under 30 percent corruption, and Figures 11(b)–11(d) separately denote the associated original face images, the reconstruction error images (the proposed algorithm with its own class and the other), and gradient magnitude images of six individuals. The first image from the left in Figure 11(c) is attained by the computation of probe face image and the training face images from the same class. And, the rest attained are from the computation when the probe face image and the training face images are from different class.

As can be seen that, due to the pixel corruption, the test image becomes blurred. Compared with the other error images, the error image extracted from its true individual still looks more distinguishable from the gradient perspective. It can be seen, from Figure 12(a), when the percentage of corrupted pixels is between 10 and 20, LRC almost correctly classifies all of the testing images. However, when

(a) (b) (c)

Figure 10: Test samples occluded on (a) nose, (b) eyes, and (c) mouth.



(a) (b)

(c)

(d)

Figure 11: (a) Test sample on 30% pixel corruption. (b) The training face images of 6 individuals. (c) Reconstruction error images of the proposed method. (d) Gradient magnitude images.

the percentage of corrupted pixels is more than 40, the recognition rate of Eigenfaces, Fisherfaces, and LRC is below 80%. The performance of SRC drops heavily beyond 50 percent corruption. We can also observe that, from 0 to 50 percent pixel corruption, our method correctly classifies all subjects. At 50 percent corruption, the proposed approach is 95.18%, while the excellent Gradientfaces is only 59.31%. Notice that none of the others achieves higher than 60% recognition rate. Even at 80 and 90 percent corruption, the proposed recognition rate is still higher than others.

For Subset 3, the faces are under even worse illumination conditions. The recognition results for different methods can be seen in Figure 12(b). We observe that our method outperforms others in this situation. For example, at 20 percent corruption, the recognition accuracy of the proposed method and LRC is 90% recognition rate. However, at 40 percent corruption, the proposed method achieves 92.00% recognition rate which outperforms SRC and LRC by a margin of 70.48% and 33.14%, respectively.

4.3. Recognition in the Presence of Block Occlusion. In this part, we test the robustness of our method to block occlusion. We use Subset 1 and Subsets 2 for training and testing separately. Different levels of contiguous occlusion from 10 to 70 percent were simulated by a square block of each test image substituting for an nonface image, that is, a baboon image. Figure 13 indicates a test image under 10 to 70 percent occlusion. The occlusion position is randomly selected in

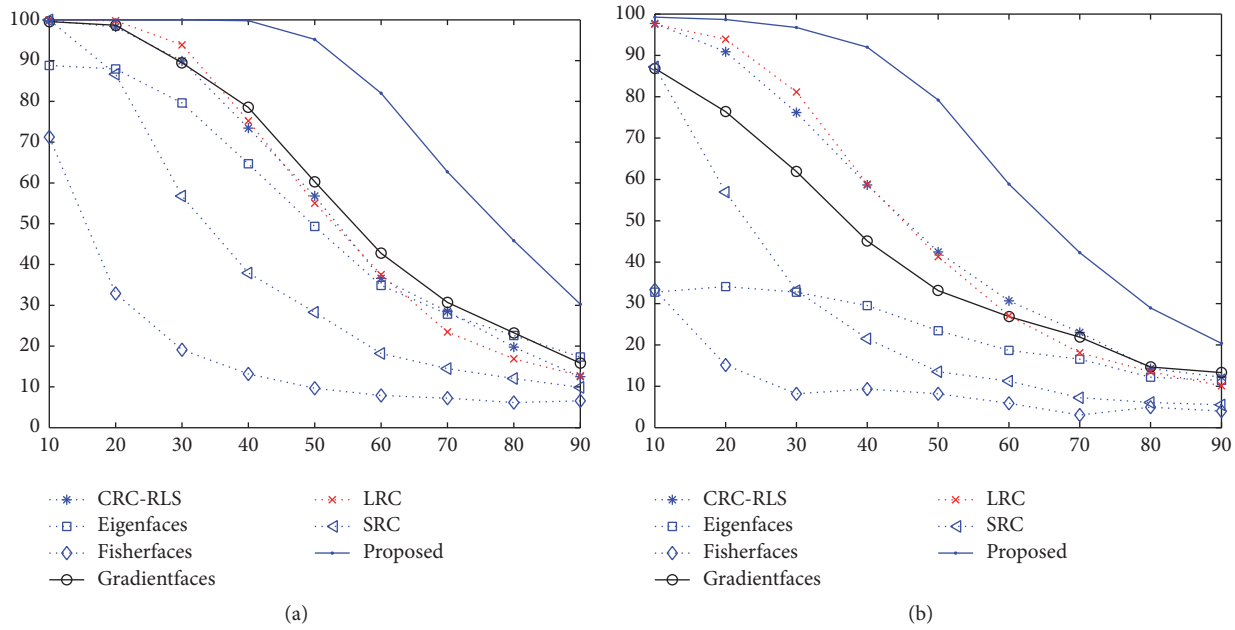


Figure 12: Random pixel corruption on Subset 2 and Subset 3 of Extended Yale B Database using Subset 1 as training sample. (a) Recognition results on Subset 2 in our reconstruction error based method. (b) Recognition results on Subset 3 in our reconstruction error based method.

Figure 13: Test sample under different degree of occlusion from 10% to 70%.

a test image and these algorithms do not know the position. Contiguous occlusion is much worse than the random pixel corruption in the face recognition algorithms.

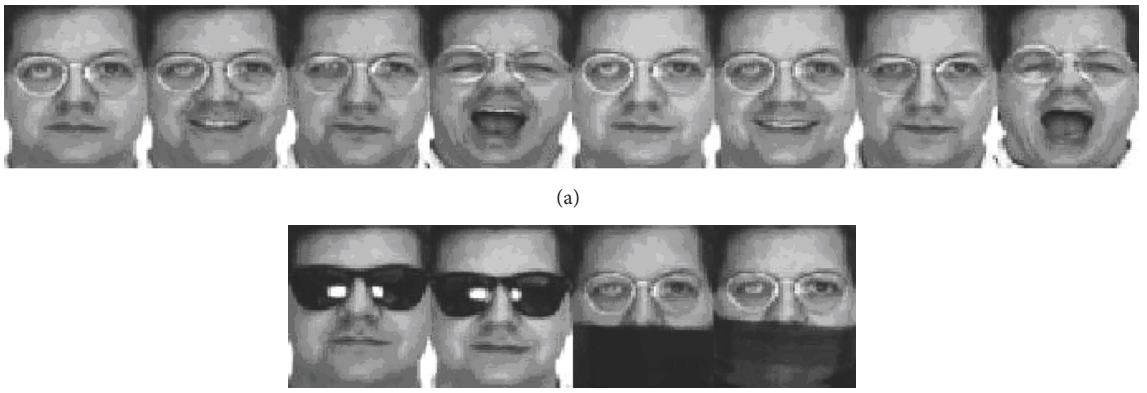
Table 3 shows that the recognition rate of SRC, LRC, and the reconstruction error based method all yield 100%. The proposed algorithms do not suppose prior knowledge about the occlusion feature. Based on an important principle of coding theory, the redundancy in the measurement is essential to detect gross errors [23]. Therefore, if a part of the image is completely corrupted by varying occlusions, identification can still be performed depending on the residual image. In this sense, our method and SRC both benefit from this principle. Table 4 illuminates the correct recognition of all compared algorithms. The proposed method significantly outperforms the other methods for different degree of occlusion. When half of a face is occluded, our method recognized over 97.80% of test subjects for Subset 3 correctly. Even at 60% occlusion, over 97.65% recognition rate can be achieved.

4.4. Recognition of Face in Disguise. In the experiment, 100 persons (including 50 males and 50 females) are selected from AR database [37]. We use 799 images (about 8 samples per face) characterized by frontal views with various expressions and without occluded for training, and another two groups of 200 faces are chosen for testing at the same time. The first testing group includes images of the faces with sunglasses, occluding approximately 20 percent in the image. The second testing group includes images of the faces wearing a scarf, occluding approximately 40 percent of the image. All face images are converted into gray scale, cropped and aligned by the centers of eyes and mouth, and then normalized with resolution 64×56, as shown in Figure 14. In addition, the robust CRC (R-CRC) [25], proposed for face recognition with occlusion, is selected for comparison.

Table 5 depicts the comparison of our method with other compared algorithms. For LRC, the recognition results, affected by the feature extraction and the number of training samples, is 66% for sunglasses and 12.5% for scarves, respectively. For the reconstruction error based method its recognition rate is 51.72% which is less than 87.00% of SRC when the occlusion is caused by scarves. However, under sunglasses occlusion, our method attains the excellent correct recognition of 94.50%, more than 28.5% of the nearest competitor LRC and better than 29% of SRC.

5. Conclusion

In the paper, we present a new method to recognize face images under illumination variations, pixel corruption, and



(a)

(b)

Figure 14: Training and test samples of a face on AR database. (a) Training samples. (b) Test samples occluded by sunglasses and scarf, separately.

Table 3: Correct recognition (%) on Subset 2 in the database of Extended Yale B under different degree of contiguous occlusion with Subset 1 as training set.

Percent occluded	0%	10%	20%	30%	40%	50%	60%	70%
Eigenfaces + NN	89.04	86.84	85.75	85.09	84.87	83.11	80.04	74.56
Fisherfaces + NN	100	100	100	100	100	99.34	100	67.98
SRC	100	100	100	100	100	100	100	100
LRC	100	100	100	100	100	100	100	100
Our method$_{error}$	100	100	100	100	100	100	100	100

Table 4: Correct recognition (%) on Subset 3 in the database of Extended Yale B under different degree of contiguous occlusion with Subset 1 as training set.

Percent occluded	0%	10%	20%	30%	40%	50%	60%	70%
Eigenfaces + NN	30.29	28.76	31.81	33.52	34.48	34.10	30.86	25.71
Fisherfaces + NN	85.52	84.95	85.14	80.95	80.38	84.19	78.29	24.95
SRC	98.86	98.10	97.33	95.43	92.76	92.19	89.52	85.52
LRC	98.67	97.52	97.33	96.57	96.19	96.57	96.00	94.10
Our method$_{error}$	99.62	97.90	98.10	97.90	97.71	97.80	97.65	96.38

Table 5: Correct recognition (%) on the database of AR in disguise.

Disguise	Sunglasses	Scarves
Eigenfaces + NN	43.00	7.50
Fisherfaces + NN	45.50	74.50
SRC	65.50	87.00
LRC	66.00	12.50
Our method$_{error}$	94.50	51.72

occlusion. The reconstruction error in constrained objective function is used for classification. Experimental results show that the reconstruction error based approach is robust to illumination variations and corruptions such as occlusion and disguise when we do not assume prior knowledge about the light sources assumption and the nature of corrupted and occluded regions. Not only is our method used in face pattern recognition, but it can be applied to other image-based object identification. Although the performance of recognition protocol based on the linear representation coefficients needs to be improved, it gives a new insight to image-based recognition. It make us know that, even in a small sample space, the class-specific samples will show unique features under some constraints. Meanwhile, we found that the other variable, the linear combination coefficients, also yields the discriminating information which is of benefit to face identification. In the future, we will improve the performance of the linear representation coefficients based method.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is sponsored by the Major Program of National Natural Science Foundation of China (no. 91438104, no.

91420102, and no. 61472053); Project Supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant no. KJ15012001); Chongqing Postdoctoral Special Funding Project (Xm2015063); and Fundamental Research Funds for the Central Universities in China (no. 106112015CDJRC161203).

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.

[2] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, and Z. Zhu, "Robust face recognition via occlusion dictionary learning," *Pattern Recognition*, vol. 47, no. 4, pp. 1559–1572, 2014.

[3] J. Huang, X. You, Y. Yuan, F. Yang, and L. Lin, "Rotation invariant iris feature extraction using Gaussian Markov random fields with non-separable wavelet," *Neurocomputing*, vol. 73, no. 4–6, pp. 883–894, 2010.

[4] Z. He, X. Li, X. You, D. Tao, and Y. Y. Tang, "Connected component model for multi-object tracking," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3698–3711, 2016.

[5] W. Zhao and R. Chellappa, *Face Processing: Advanced Modeling and Methods*, Academic Press, 2011.

[6] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: a survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.

[7] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma, "Face recognition with contiguous occlusion using Markov Random Fields," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 1050–1057, IEEE, October 2009.

[8] Q.-X. Gao, L. Zhang, and D. Zhang, "Face recognition using FLDA with single training image per person," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 726–734, 2008.

[9] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2599–2606, 2009.

[10] Y. Wang, L. Zhang, Z. Liu et al., "Face relighting from a single image under arbitrary unknown lighting conditions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1968–1984, 2009.

[11] X. Xie, W.-S. Zheng, J. Lai, P. C. Yuen, and C. Y. Suen, "Normalization of face illumination based on large- and small-scale features," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1807–1821, 2011.

[12] O. Arandjelović, "Gradient edge map features for frontal face recognition under extreme illumination changes," in *Proceedings of the 2012 23rd British Machine Vision Conference (BMVC '12)*, pp. 1–11, September 2012.

[13] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 748–763, 2002.

[14] S. Fidler, D. Skočaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 337–350, 2006.

[15] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[16] D. Xu, S. Yan, D. Tao, S. Lin, and H.-J. Zhang, "Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2811–2821, 2007.

[17] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[18] H. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 157–165, 2006.

[19] J. Liu, S. Chen, X. Tan, and D. Zhang, "Comments on 'Efficient and robust feature extraction by maximum margin criterion'," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1862–1864, 2007.

[20] J. Kim, J. Choi, J. Yi, and M. Turk, "Effective representation using ICA for face recognition robust to local distortion and partial occlusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1977–1981, 2005.

[21] M. Bilginer Gülmezoğlu, V. Dzhafarov, and A. Barkana, "The common vector approach and its relation to principal component analysis," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 655–662, 2001.

[22] 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.

[23] 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.

[24] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: robust alignment and illumination by sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372–386, 2012.

[25] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, "Collaborative representation based classification for face recognition," https://arxiv.org/abs/1204.2358.

[26] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

[27] S. Ding, N. Zhang, X. Xu, L. Guo, and J. Zhang, "Deep extreme learning machine and its application in EEG classification," *Mathematical Problems in Engineering*, vol. 2015, Article ID 129021, 11 pages, 2015.

[28] H. Luo, Y. Y. Tang, and L. Yang, "Subspace learning via local probability distribution for hyperspectral image classification," *Mathematical Problems in Engineering*, vol. 2015, Article ID 145136, 17 pages, 2015.

[29] Z. Shi, "A weighted block dictionary learning algorithm for classification," *Mathematical Problems in Engineering*, vol. 2016, Article ID 3824027, 15 pages, 2016.

[30] W. Zhao and R. Chellappa, *Robust Face Recognition Using Symmetric Shape-From-Shading*, Computer Vision Laboratory, Center for Automation Research, University of Maryland, Maryland, Md, USA, 1999.

[31] 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," *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, vol. 23, no. 6, pp. 643–660, 2001.

[32] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[33] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal L_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[34] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[35] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[37] A. Martinez, "The ar face database," CVC Technical Report Issue: 24, CVC, 1998.