A New Discriminative Sparse Representation Method for Robust Face Recognition via l_2 Regularization

Yong Xu, *Senior Member, IEEE*, Zuofeng Zhong, Jian Yang, *Member, IEEE*, Jane You, and David Zhang, *Fellow, IEEE*

***Abstract*— Sparse representation has shown an attractive performance in a number of applications. However, the available sparse representation methods still suffer from some problems, and it is necessary to design more efficient methods. Particularly, to design a computationally inexpensive, easily solvable, and robust sparse representation method is a significant task. In this paper, we explore the issue of designing the simple, robust, and powerfully efficient sparse representation methods for image classification. The contributions of this paper are as follows. First, a novel discriminative sparse representation method is proposed and its noticeable performance in image classification is demonstrated by the experimental results. More importantly, the proposed method outperforms the existing state-of-the-art sparse representation methods. Second, the proposed method is not only very computationally efficient but also has an intuitive and easily understandable idea. It exploits a simple algorithm to obtain a closed-form solution and discriminative representation of the test sample. Third, the feasibility, computational efficiency, and remarkable classification accuracy of the proposed l_2 regularization-based representation are comprehensively shown by extensive experiments and analysis. The code of the proposed method is available at http://www.yongxu.org/lunwen.html.**

***Index Terms*— Efficient computation, face recognition, l_2 regularization, sparse representation.**

I. Introduction

SPARSE representation has drawn much attention since it was proposed in the pattern recognition and machine learning communities [1]–[3]. Sparse representation has achieved a noticeable performance in image classification [4]–[7], image super-resolution [8], [9], image denoising [10], and image alignment [11].

A number of variants of the sparse representation methods have been proposed in recent years. The taxonomy of the sparse representation methods for image classification can be considered in different ways. In terms of the atoms used, the sparse representation methods can be categorized into two kinds, the Naive training sample-based and dictionary-based methods. The Naive training sample-based sparse representation directly exploits the original training samples as the atoms to produce the representation of the test sample, whereas the dictionary-based sparse representation first uses an algorithm and the original training samples to obtain a dictionary, and then exploits this dictionary as the atoms to lead to the representation of the test sample [5], [12]–[17], [44].

It is often that some constraints are imposed on the sparse representation methods. The most widely used constraints include l_1 regularization [1], [2], l_2 regularization [18]–[20] and l_{21} regularization [21], [22]. The l_2 regularization-based representation method has distinctive advantages. One advantage is that it almost has a closed-form solution.

In this paper, we propose a novel idea to design the sparse representation method by enhancing the distinctiveness of different classes. This is very beneficial to produce higher classification accuracy. A simple, but competent objective function is proposed, and a very computationally efficient algorithm is designed. Moreover, the proposed method is based on the mathematically tractable l_2 regularization rather than the l_1 or l_{21} regularization. The representation coefficients obtained using the l_2 regularization-based representation method are not as sparse as those obtained using the l_1 regularization-based representation method. The proposed method is reasonable and compatible with the nature of sparse representation due to the following factors. First of all, the sparse representation exploits the residuals of different classes, i.e., class-specific residuals as the distances of these classes with respect to the test sample. The proposed method enables different classes to be more discriminative. As a result, it can obtain a great class difference, which is helpful for the representation methods to obtain discriminative class-specific residuals and to better recognize the test sample.

The other parts of the paper are organized as follows. Section II presents related works. Section III describes the proposed method including its advantages and rationale. Section IV presents a comparison between the proposed

Manuscript received September 30, 2015; revised January 11, 2016, April 20, 2016, and June 2, 2016; accepted June 8, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61370163 and Grant 61300032, in part by the Shenzhen Municipal Science and Technology Innovation Council under Grant JCYJ20150330155220591 and Grant JCYJ20140904154645958, in part by the Hong Kong Government through the General Research Fund Scheme under Grant 152202/14E, and in part by the Research Grant through The Hong Kong Polytechnic University under Grant G-YN39. *(Corresponding author: Yong Xu.)*

Y. Xu is with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China, and also with the Key Laboratory of Network Oriented Intelligent Computation, Shenzhen 518055, China (e-mail: yongxu@ymail.com).

Z. Zhong is with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055,China (e-mail: zfzhong2010@gmail.com).

J. Yang is with the School of Computer Science and Technology, Nanjing University of Science of Science and Technology, Nanjing 210094, China (e-mail: csjyang@mail.njust.edu.cn).

J. You and D. Zhang are with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (e-mail: csyjia@comp.polyu.edu.hk; csdzhang@comp.polyu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2016.2580572

method and other methods. Sections V and VI offer the experimental results and conclusion, respectively.

II. Related Works

Let L be the total number of all classes. If each class has m training samples, then there are $N(N = mL)$ training samples in total. We use the column vectors $x_{m(c-1)+1}, x_{m(c-1)+2}, \ldots, x_{cm}$ to stand for the m training samples of the cth $(c = 1, 2, \ldots, L)$ class, respectively. We define training sample matrix $X = [X_1, K, \ldots, X_c, \ldots, X_L] = [x_1, K, \ldots, x_{m(c-1)+1}, \ldots, x_{mc}, \ldots, x_N]$, where $X_c = [x_{m(c-1)+1}, x_{m(c-1)+2}, \ldots, x_{mc}]$, and let column vector y denote the test sample.

A. l_2 Regularization-Based Representation

l_2 regularization-based representation is a kind of computationally efficient method. Collaborative representation classification (CRC) is a typical example of the l_2 regularization-based representation. In addition to CRC, a small number of other l_2 regularization-based representation methods have also been proposed. For example, both robust regression for classification [23] and two-phase test sample sparse representation [20], [24] are l_2 regularization-based representation methods. If Naive l_2 regularization is directly applied, it seems that the sparse property will not be well guaranteed [25]. However, as shown in [18], [20], and [24], by exploiting a special means, l_2 regularization can also obtain satisfactory sparse property. Especially, the elaborated algorithms of l_2 regularization-based representation, such as the ones designed in [20] and [24], can obtain quite sparse representation coefficients, and flexibly control the extent of sparsity.

In this section, we briefly introduce the CRC as an example of the l_2 regularization-based representation methods. CRC first produces a representation of the test sample via a linear combination of all the training samples. In other words, CRC first assumes that the following equation is approximately satisfied and solves it using the ridge regression:

$$y = \sum_{i=1}^{N} a_i x_i. \tag{1}$$

Equation (1) can be rewritten as

$$y = XA, \quad A = [a_1, \ldots, a_N]^T. \tag{2}$$

Here, we refer to a_i $(i = 1, 2, \ldots, N)$ as the coefficients.

CRC solves (2) using $\bar{A} = (X^T X + \mu I)^{-1} X^T y$, where μ is a small positive constant and I is the identity matrix [19]. Hereafter, $\bar{A}$ is referred to as the representation coefficients, i.e., solution. If some of the representation coefficients are zero or close to zero, we can refer to the solution as sparse representation coefficients. Let $\bar{A}_j = [\bar{a}_{m(j-1)+1}, \ldots, \bar{a}_{mj}]^T$, where $1 \le j \le L$ and $\bar{a}_{mj}$ stands for the mjth entry of $\bar{A}$. We refer to $\bar{A}_j$ as the coefficient vector of the jth class. The classification rule of CRC is as follows: the dissimilarity between the test sample y and the jth class is defined as $d_j = (||y - X_j \bar{A}_j||/||\bar{A}_j||)$. If $k = \arg\min_j d_j$, then the test sample y is assigned to the kth class.

B. l_1 Regularization-Based Representation

l_1 regularization-based representation solves the following problem [2]:

$$\min ||A||_1 \quad \text{s.t. } y = XA \tag{3}$$

or

$$\min ||A||_1 \quad \text{s.t. } ||y - XA||_2 < \varepsilon \tag{4}$$

where ε stands for a small positive constant. l_1 regularization-based representation should be implemented by using an iterative algoriTh. Important and noticeable l_1 regularization-based representation algorithms include l_1-regularized least squares (L1LS) [26], fast iterative shrinkage and thresholding algorithm [27], [32], homotopy and augmented Lagrangian [28], [29], and orthogonal matching pursuit [30].

The classification procedure of l_1 regularization-based representation is as follows. Let $\bar{A}$ be the solution of l_1 regularization-based representation. $\bar{A}_j$ is also the coefficient vector of the jth class. The dissimilarity between the test sample y and the jth class is defined as $d_j = ||y - X_j \bar{A}_j||$. If $k = \arg\min_j d_j$, then the test sample y is assigned to the kth class.

III. Proposed Method

Suppose that the test sample y and each training sample $x_{m(i-1)+j}$ $(i = 1, \ldots, L,\ j = 1, K, \ldots, m)$ are the D-dimensional column vectors. Because there are L classes and N training samples in total, X is a D by N matrix.

A. Description of the Method

The objective function of the proposed method is defined as

$$\min_B \|y - XB\|^2 + \gamma \sum_{i=1}^{L} \sum_{j=1}^{L} \|X_i B_i + X_j B_j\|^2 \tag{5}$$

where γ denotes a positive constant, and is used to balance the effect of both terms in the objective function of the proposed method. When it is set to a proper value, these two terms in the objective function can have good effect. The relationship among B_i, B_j, and B is defined as follows. If $B = [b_1, b_2, \ldots, b_N]^T$, then $B_i = [b_{m(i-1)+1}, b_{m(i-1)+2}, \ldots, b_{mi}]^T$, $B_j = [b_{m(j-1)+1}, b_{m(j-1)+2}, \ldots, b_{mj}]^T$, and $B = [B_1, B_2, \ldots, B_L]^T$.

It can be verified that the objective function in (5) is convex and differentiable, and thus the optimal solution of (5) is just the stationary point of the objective function. The derivative of the objective function is computed as follows. First

$$\frac{d}{dB}\|y - XB\|^2 = -2X^T(y - XB). \tag{6}$$

Then, we need to determine $(d/dB)(\gamma \sum_{i=1}^{L} \sum_{j=1}^{L} \|X_i B_i + X_j B_j\|^2)$.

Since $f(B) = \gamma \sum_{i=1}^{L} \sum_{j=1}^{L} \|X_i B_i + X_j B_j\|^2$ dose not explicitly contain B, we first seek partial derivatives $(\partial f/\partial B_k)$, and then exploit all $(\partial f/\partial B_k)$ $(k = 1, \ldots, L)$ to obtain (df/dB).

$f(B)$ is the sum of L^2 terms of which only $2L - 1$ terms are dependent on B_k. When determining $(\partial f/\partial B_k)$, we only

need to calculate the partial derivatives of these $2L-1$ terms with respect to B_k, which is presented in detail as follows:

$$\begin{aligned} & f(B) \\ &= \gamma \left(\sum_{\substack{j=1,\ldots,L \\ j\neq k}} \|X_k B_k + X_j B_j\|^2 + \sum_{\substack{i=1,\ldots,L \\ i\neq k}} \|X_i B_i + X_k B_k\|^2 \right. \\ &\qquad \left. + \sum_{\substack{i=1,\ldots,L \\ i\neq k}} \sum_{\substack{j=1,\ldots,L \\ j\neq k}} \|X_i B_i + X_j B_j\|^2 \right) \\ &= \gamma \left(2 \sum_{\substack{j=1,\ldots,L \\ j\neq k}} \|X_k B_k + X_j B_j\|^2 \right. \\ &\qquad \left. + \sum_{\substack{i=1,\ldots,L \\ i\neq k}} \sum_{\substack{j=1,\ldots,L \\ j\neq k}} \|X_i B_i + X_j B_j\|^2 \right). \end{aligned} \tag{7}$$

Hence, the partial derivative over B_k of $f(B)$ is

$$\begin{aligned} \frac{\partial f}{\partial B_k} &= \frac{\partial}{\partial B_k} \left(\gamma \sum_{i=1}^{L} \sum_{j=1}^{L} \|X_i B_i + X_j B_j\|^2 \right) \\ &= \gamma \frac{\partial}{\partial B_k} \left(2 \sum_{\substack{j=1,\ldots,L \\ j\neq k}} \|X_k B_k + X_j B_j\|^2 \right. \\ &\qquad \left. + \sum_{\substack{i=1,\ldots,L \\ i\neq k}} \sum_{\substack{j=1,\ldots,L \\ j\neq k}} \left\|X_i B_i + X_j B_j\right\|^2 \right) \end{aligned} \tag{8}$$

$$\begin{aligned} \frac{\partial f}{\partial B_k} &= 2\gamma \sum_{\substack{j=1,\ldots,L \\ j\neq k}} \frac{\partial}{\partial B_k} \|X_k B_k + X_j B_j\|^2 \\ &= 2\gamma \sum_{\substack{j=1,\ldots,L \\ j\neq k}} 2X_k^T (X_k B_k + X_j B_j) \\ &= 4\gamma X_k^T \left((L-1) X_k B_k + \sum_{\substack{j=1,\ldots,L \\ j\neq k}} X_j B_j \right) \\ &= 4\gamma X_k^T \left(L X_k B_k + \sum_{j=1}^{L} X_j B_j \right) \\ &= 4\gamma X_k^T (L X_k B_k + XB). \end{aligned} \tag{9}$$

Thus, the derivative (df/dB) is

$$\begin{aligned} \frac{df}{dB} &= \begin{pmatrix} \frac{\partial f}{\partial B_1} \\ \vdots \\ \frac{\partial f}{\partial B_L} \end{pmatrix} = \begin{pmatrix} 4\gamma X_1^T (L X_1 B_1 + XB) \\ \vdots \\ 4\gamma X_L^T (L X_L B_L + XB) \end{pmatrix} \\ &= 4\gamma L \begin{pmatrix} X_1^T X_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & X_L^T X_L \end{pmatrix} B + 4\gamma X^T XB. \end{aligned} \tag{10}$$

Algorithm 1 l_2 Regularization Based Discriminative Sparse Representation Algorithm

Input: γ, y, $X = [X_1, \ldots, X_L]$.

Output: d_i, p, $B = [B_1, \ldots, B_L]^T$

1. B is calculated using
$$B = \left((1+2\gamma) X^T X + 2\gamma LM\right)^{-1} X^T y.$$
2. The distance between the test sample y and the i-th class training samples is obtained using class-specific residual
$$d_i = ||X_i B_i - y||_2^2, \quad i = 1, \ldots, L.$$
3. Test sample y is classified to the p-th class, if $p = \arg\min_i d_i$.

Let

$$M = \begin{pmatrix} X_1^T X_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & X_L^T X_L \end{pmatrix}$$

then

$$\frac{df}{dB} = 4\gamma (LMB + X^T XB).$$

In conclusion, let $g = \|y - XB\|^2 + \gamma \sum_{i=1}^{L} \sum_{j=1}^{L} \|X_i B_i + X_j B_j\|^2$, then we have $(dg/dB) = -2X^T (y - XB) + 4\gamma (LMB + X^T XB)$. Thus, the stationary point of g is obtained under the condition $(dg/dB) = 0$, which means $((1+2\gamma) X^T X + 2\gamma LM) B = X^T y$. Finally the optimal solution of (5) is

$$B = ((1+2\gamma) X^T X + 2\gamma LM)^{-1} X^T y. \tag{11}$$

The proposed method then uses the same classification procedure as the conventional sparse representation methods. In order to provide a full description, the algorithm of the proposed method is presented in Algorithm 1.

IV. Insight Into the Proposed Method

In this section, we present the advantage, rationale, and computational complexity of the proposed method.

A. Advantages and Computational Complexity of the Proposed Method

The first advantage of the proposed method is that it has a closed-form solution. This is because the proposed method is based on the l_2 regularization rather than the l_1 or l_{21} regularization. For the same reason, the proposed method is mathematically tractable and computationally efficient. It is notable that, as is shown later, the proposed method achieves much higher accuracy than CRC, the typical example of the l_2 regularization-based representation methods. Moreover, it is that the representation coefficients obtained using our method and the l_1 regularization-based representation methods have no essential difference. The representation coefficients obtained using our method and an l_1 regularization-based representation method, and the L1LS method [26] are shown

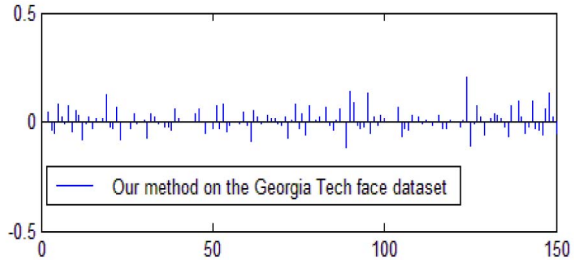

Fig. 1. Representation coefficients on the last test sample of the GT face data set (There are 50 subjects, and each subject has 15 images) obtained using our method. The first three face images of each subject were used for training, and the remaining images were used for testing. The horizontal axis shows the number of training sample, and the vertical axis shows the representation coefficient corresponding to each training sample. The horizontal vertical axes in Figs. 2–4 play the same roles.

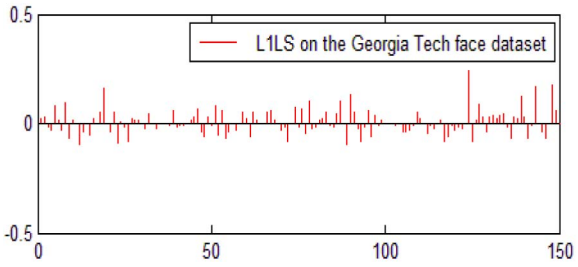

Fig. 2. Representation coefficients on the last test sample of the GT face data set obtained using the L1LS method issued via [25]. The first three face images of each subject were used for training, and the remaining images were used for testing.

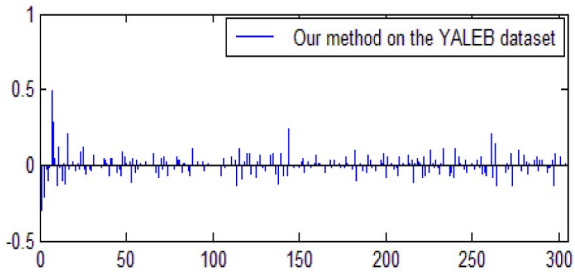

Fig. 3. Representation coefficients on the second test sample obtained using our method on the Extended YaleB face data set (There are 38 subjects, and each subject has 64 images). The first eight face images of each subject were used for training, and the remaining images were used for testing.

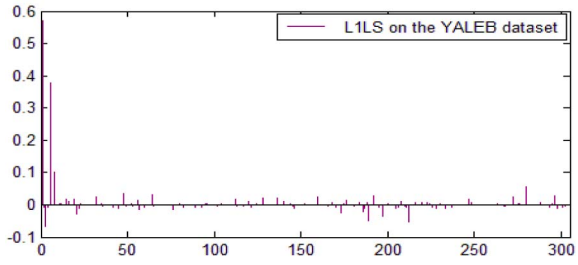

Fig. 4. Representation coefficients on the second test sample obtained using the L1LS algorithm issued via [25] on the Extended YaleB face data set. The first eight face images of each subject were used for training, and the remaining images were used for testing.

in Figs. 1–4. It can be seen from Figs. 1–4 that the representation coefficients of these two methods have somewhat similar distributions.

In order to quantify the sparsity, we conducted a special experiment. We used a widely applied sparsity metric

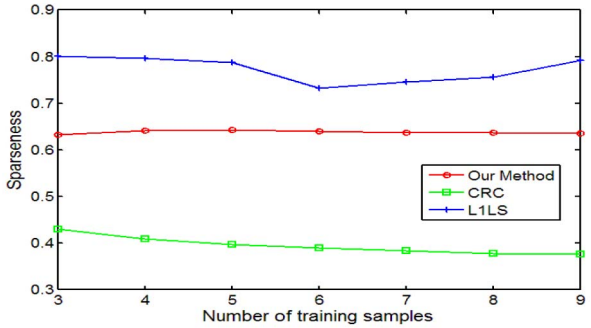

Fig. 5. Comparison of average sparseness among our method, the CRC method, and the L1LS method on the GT data set. The horizontal axis shows the number of training sample per class, and the vertical axis shows average sparseness of the representation coefficients corresponding to the test samples.

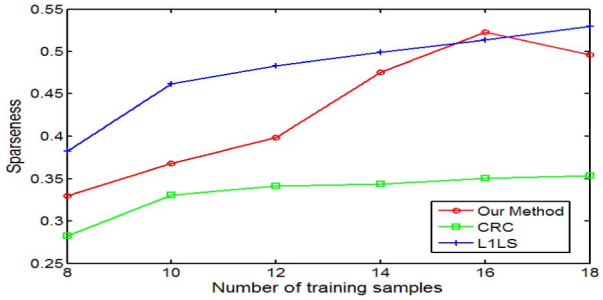

Fig. 6. Comparison of average sparseness among our method, the CRC method, and the L1LS method on the Extended YaleB data set. The horizontal axis shows the number of training sample per class, and the vertical axis shows average sparseness of the representation coefficients corresponding to the test samples.

to compare the proposed method, the CRC method, and the L1LS method. For a solution vector, i.e., the representation coefficients in the form of $x = [x_1, x_2, \ldots, x_N]$, the sparsity metric is defined as sparseness $(x) = (\sqrt{N} - (\sum |x_i|)/(\sum x_i^2)^{1/2}/\sqrt{N} - 1)$ [31]. From the formula, we know that the larger the value of sparseness, the sparser the vector. Figs. 5 and 6 present the comparison results on the GT data set and the Extended YaleB data set, respectively. They denote average sparseness for all the test samples under the condition that a different number of face images per class are used as the training samples. It can be seen that our method achieves better sparsity than the CRC method on both data sets. However, the proposed method is less sparse than the L1LS method on the GT data set. On the Extended YaleB data set, the sparseness of the L1LS method is better than that of our method under the condition that 8, 10, 12, 14, and 18 face images per class are used as the training samples. Thus, we believe that our method is not always less sparse than L1LS, which obtains better sparsity under some conditions.

The computational complexity of the proposed method is described as follows. The main computations are the matrix operations in (11). The computational complexity of $X^T X$ in (10) is $O(DN^2)$. Because of $X = [X_1, \ldots, X_L]$, we have

$$X^T X = \begin{pmatrix} X_1^T X_1 & \cdots & X_1^T X_L \\ \vdots & \ddots & \vdots \\ X_L^T X_1 & \cdots & X_L^T X_L \end{pmatrix}.$$

As a result, after we calculate X^TX, we can directly obtain M, and no extra computational complexity is needed. Let $H = (1+2\gamma)X^TX + 2\gamma LM$, then H is N by N matrix. Thus, the computational complexity of H^{-1} is $O(N^3)$. The computational complexity of $P = H^{-1}X^T$ is also $O(DN^2)$. The computational complexity of $Q = Py$ is also $O(DN)$. It should be pointed out that in our method $P = H^{-1}X^T$ needs to be calculated only once, whereas $Q = Py$ should be implemented for every test sample. As a result, if there are R test samples in total, then the computational complexity of (11) will be $O(DN^2+N^3+DNR)$. Actually, this is similar to that of CRC and is much less than those of the l_1 regularization-based sparse representation methods.

We analyzed the computational complexity of eight methods in our comparison experiment including CRC [19], L1LS [26], FISTA [32], homotopy [28], the dual augmented lagrangian method (DALM) [28], improvement to the nearest neighbor classifier (INNC) [43], a fusion classification method (FCM) proposed in [44], and linear discriminant analysis (LDA) [45]. We assumed that there are R test samples. When we used CRC to perform classification of all the R test samples, the main computational load of CRC was in the calculation of matrix $P = (X^TX + \mu I)^{-1}X^T$ whose complexity is $O(DN^2+N^3)$. Thus, for the R test samples, CRC has a computational complexity of $O(DN^2+N^3+DNR)$, which is similar to that of our method. Because the L1LS, FISTA, homotopy, and DALM are different algorithms of SRC, their computational complexity is also different. For L1LS, FISTA, homotopy, and DALM, the computational complexity for classification of all the R test samples is at least $O(RTDN)$, $O(RDN^2+RDN)$, $O(RN^2+RDN)$, and $O(RTD^2+RTND)$, respectively, where T is the iteration number. The computational complexity of both the INNC and FCM methods are $O(DN^2+N^3+RDN)$. The LDA method mainly needs to compute the within-class and between-class scatter matrices, and the computational complexity is $O(D^3+RD^2)$. The computational superiority of our method is twofold. First, it has a closed-form solution, and no iterative calculation is needed. Second, when there are a large number of test samples, our method has a lower computational complexity than L1LS, FISTA, homotopy, and DALM.

B. Rationale of the Proposed Method

The proposed method is able to enhance discriminant representations of the test sample generated from different classes, which is beneficial for correct classification of the test sample. Differing from our work, previous methods that have a similar effect mainly focused on obtaining different dictionaries for different classes via dictionary learning rather than directly enhancing the distinction of the original training samples. For example, Ramirez *et al.* [33] generated the dictionaries associated with different classes with the goal that the independence is as great as possible. Yang *et al.* [13] produced a discriminative dictionary for every class. Discriminative dictionary learning and discriminative sparse representations were also applied to image analysis, clustering analysis [34], motion tracking [35], [36], and image segmentation [37]. Compared with these methods, our method is simpler and easier to implement.

The proposed method has the following rationale. First of all, we refer to X_iB_i as the representation result of the test sample obtained using the training samples of the ith class. The first term of the objective function of our method is the same as the conventional sparse representation methods and aims to obtain the minimum residual, the reasonability of which is easy to understand. In order to interpret the second term of the objective function, we first rewrite $\|X_iB_i + X_jB_j\|^2$ as $\|X_iB_i + X_jB_j\|^2 = \|X_iB_i\|^2 + \|X_jB_j\|^2 + 2(X_iB_i)^T(X_jB_j)$. Thus, to minimize $\|X_iB_i + X_jB_j\|^2$ is equivalent to simultaneously minimize $(X_iB_i)^T(X_jB_j)$ and $\|X_iB_i\|^2$, $i = 1, \ldots, L$. Minimization of $(X_iB_i)^T(X_jB_j)$ means that the representation results of the ith and jth classes have the lowest correlation, which enables the representation results of different classes to be very discriminative. Thus, the second term of the objective function is able to decorrelate the representation results of different classes. The decorrelation effect for different classes can only be achieved by minimizing the sum of $(X_iB_i)^T(X_jB_j)$, not any individual terms. As a result, the method will more easily distinguish the class really nearest to the test sample from the other classes and will more easily correctly classify the test sample. Minimization of $\|X_iB_i\|^2 (i = 1, \ldots, L)$ implies that the representation result of each class has a small norm. In addition, similar to other representation methods, in our method, the training samples of different classes also compete in representing the test sample. In other words, since the test sample is expressed as a weighted sum (i.e., a linear combination) of the training samples of all the classes, each class has a contribution to representing the test sample. Competition means that when training samples of a class have a great contribution to representing the test sample, other classes have relatively less contribution.

Both the proposed method and CRC belong to l_2 regularization-based representation. The proposed method is as computationally very efficient as CRC, but has an elaborated design and goal. As presented earlier, the objective function and motivation of the proposed method are distinctive in comparison with CRC. Moreover, it can be seen from the solutions of the proposed method and CRC that they also have the following difference on the regularization term, which is important in obtaining numerically stable solutions. As shown in Section II-A, the regularization term in the solution of CRC is μI. In other words, the effect of the regularization term of CRC is to add only small positive constants to the diagonals of X^TX.

In the algorithm of our proposed method, the regularization term is

$$2\gamma LM = 2\gamma L \begin{pmatrix} X_1^TX_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & X_L^TX_L \end{pmatrix}.$$

It has the following effect: it not only regularizes the diagonals of X^TX but also regularizes the $m \times m$ diagonal regions

Fig. 7. Some face images from the ORL face data set. The face images shown in the first, second, and third rows are from three different subjects.

of $X^T X$. This takes into account the correlation between the samples from the same class.

V. Experimental Results

In this section, we used several data sets to test and verify the effectiveness of the proposed method. The data sets used include the FERET [38], ORL [39], Georgia Tech (GT) [40], Extended YaleB [41], and CMU Multi-PIE [46] face data sets. We also tested CRC [19], L1LS [26], FISTA [32], homotopy [28], DALM [29], INNC [43], and an FCM proposed in [44] for experimental comparison. LDA [45] integrated with the nearest neighbor classifier was also tested. All experiments showed that our proposed method obtained the highest accuracy. We chose the optimal value of parameter γ for each data set from eight candidate values: 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10, and 100.

A. Experiments on the ORL Face Data Set

The ORL data set [39] includes 400 face images taken from 40 subjects, each provided with ten face images. For some subjects, the images were taken at different times, with varying lighting, facial expressions (open/closed eyes, smiling/not smiling), and facial details (glasses/no glasses). Each image was resized to an image with one half of the original size by using the down-sampling algorithm. Fig. 7 shows some face images from this data set. The parameter γ in our method was set to 0.001.The first 2–6 face images of each subject were used as training samples and the remaining face images were used as test samples. The experimental results are presented in Table I.

B. Experiments on the Extended YaleB Face Data Set

The Extended YaleB [41] face data set contains 2414 single frontal facial images of 38 individuals. These images were captured under various controlled lighting conditions. The size of an image was 192 168 pixels. In our experiments, all images were cropped and resized to 84 96 pixels. Fig. 8 shows some face images from the Extended YaleB face data set. The first 8, 10, 12, 14, 16, and 18 face images of each subject were treated as original training samples and treated the remaining face images that were viewed as testing samples. The parameter γ was set to 0.001. The experimental results have presented in Table II.

TABLE I

Classification Accuracies of Different Methods on the ORL Face Data Set

Number of training samples per subject	2	3	4	5	6
Our method	**86.88**	**89.64**	**94.17**	**94.50**	**95.00**
CRC	83.44	86.08	89.17	88.50	91.87
L1LS	86.25	88.57	92.08	92.50	93.75
FISTA	82.50	84.64	88.75	88.50	92.50
INNC	81.56	82.14	87.50	86.50	86.25
FCM	84.69	86.43	89.58	90.50	92.50
LDA	74.69	85.31	89.06	89.06	93.13
Homotopy	79.69	82.50	87.50	88.50	93.75
DALM	86.88	89.64	92.08	90.00	94.37

Fig. 8. Some face images from the Extended YaleB face data set. The face images shown in the first and second rows are from two different subjects.

Fig. 9. Some face images from the GT face data set. The face images shown in the first, second, and third rows are from three different subjects.

C. Experiments on the Georgia Tech Face Data Set

There are 750 available face images (50 subjects and each has 15 face images) in the GT face data set [40]. These face images show frontal and/or tilted faces with different facial expressions, lighting conditions and scale. In our experiments, all images in the data set were manually cropped and resized to 30×40. Fig. 9 shows some cropped face images from the GT face data set. We converted every image into gray level images for training and testing. The first 3–9 face images of each subject were used as original training samples, and the remaining face images were treated as testing samples. The parameter γ was set to 1. The experimental results are shown in Table III.

D. Experiments on a Subset of the FERET Face Data Set

This subset used is from the well-known FERET face data set [38], which contains 1400 face images from 200 subjects, each provided with seven different face images. This subset is composed of images in the original FERET face data set

TABLE II
CLASSIFICATION ACCURACIES OF DIFFERENT METHODS ON THE EXTENDED YaleB FACE DATA SET

Number of training samples per subject	8	10	12	14	16	18
Our method	**69.55**	**78.90**	**80.31**	**80.79**	**82.18**	**82.61**
CRC	62.73	67.54	70.45	71.53	73.14	74.43
L1LS	66.54	73.25	75.35	76.32	76.70	78.15
FISTA	42.29	42.35	41.24	39.89	39.75	40.96
INNC	65.27	72.47	74.34	74.53	75.99	76.89
FCM	59.87	65.20	67.46	67.89	69.85	71.34
LDA	62.22	65.51	66.49	67.67	68.70	71.19
Homotopy	44.78	43.86	43.78	44.63	46.55	46.00
DALM	53.85	59.45	64.17	65.84	68.31	69.16

TABLE III
CLASSIFICATION ACCURACIES OF DIFFERENT METHODS ON THE GT FACE DATA SET

Number of training samples per subject	3	4	5	6	7	8	9
Our method	**53.67**	56.55	**61.80**	**70.67**	**73.25**	**75.14**	**78.00**
CRC	45.33	47.09	48.80	55.56	58.50	59.43	61.00
L1LS	51.00	57.27	61.60	68.67	71.25	73.71	75.00
FISTA	48.83	51.09	52.20	61.11	64.25	66.29	65.33
INNC	40.50	40.54	42.80	46.67	48.25	48.29	52.67
FCM	47.67	50.00	52.20	57.78	62.00	63.43	68.67
LDA	42.00	48.67	52.83	61.00	66.00	73.67	77.83
Homotopy	45.50	49.09	51.40	59.33	62.75	65.71	68.00
DALM	52.50	**57.45**	61.20	69.56	72.50	75.14	76.33

Fig. 10. Some face images from the FERET face data set. The face images shown in the first, second, and third rows are from three different subjects.

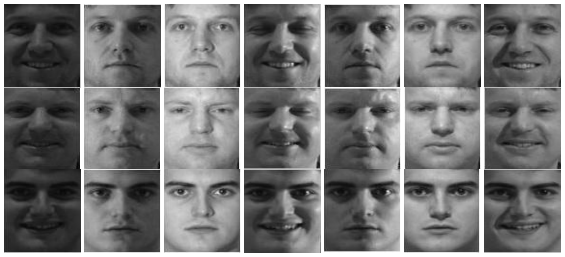

Fig. 11. Some face images from the CMU Multi-PIE face data set. The face images shown in the first, second, and third rows are from three different subjects. These images show effects on facial appearance under different illuminations.

whose names are marked with two-character strings: ba, bj, bk, be, bf, bd, and bg. Fig. 10 shows some of these face images. Every face image was resized to a 40×40 image. The first 2–5 face images of each subject and the remaining face images were used as the training samples and the test samples. The parameter γ was set to 100. The experimental results are presented in Table IV.

E. Experiments on the CMU Multi-PIE Face Data Set

In this section, we evaluated the performance of our method on the CMU Multi-PIE face data set [46]. The CMU Multi-PIE face data set is composed of a face image of 337 persons with variations of poses, expressions, and illuminations. Fig. 11 shows some face images from this data set. We select a subset composed of 249 persons under 20 different illumination conditions with a frontal pose and 7 different illumination conditions with a smile expression. All the images are cropped and resized to 40×30 pixels. We choose face images corresponding to the first 3, 5, 7, 9 illuminations from the 20 illuminations and only one image from 7 smiling images as the training samples and use remaining images as the testing samples. The parameter γ is set to 0.00001. Table V lists the classification accuracy on four testing sets obtained

TABLE IV
CLASSIFICATION ACCURACIES OF DIFFERENT METHODS ON THE FERET FACE DATA SET

Number of training samples per subject	2	3	4	5
Our method	**65.80**	**61.62**	**79.00**	80.50
CRC	58.40	44.37	55.33	68.50
L1LS	65.40	59.13	76.00	**82.00**
FISTA	47.80	38.75	49.00	58.00
INNC	58.30	50.50	54.00	68.75
FCM	57.20	47.37	56.50	68.25
LDA	40.60	47.80	65.60	81.60
Homotopy	59.10	54.00	72.50	77.25
DALM	64.90	59.75	75.17	80.00

TABLE V
CLASSIFICATION ACCURACIES OF DIFFERENT METHODS ON THE SUBSET OF THE CMU MULTI-PIE FACE DATA SET

Number of training samples per subject	4	6	8	10
Our method	**93.68**	**99.18**	**99.85**	**99.86**
CRC	90.13	96.63	98.86	99.46
L1LS	92.79	99.08	99.66	99.81
FISTA	46.74	50.01	57.26	71.16
INNC	82.05	95.62	98.82	99.50
FCM	90.34	95.47	97.36	98.32
LDA	89.80	95.09	98.01	99.31
Homotopy	55.49	56.55	63.03	71.25
DALM	68.05	67.62	72.27	84.67

using different methods. From the results, it can be seen that our method obtains better classification accuracy than other methods. In other words, our method is more robust to the variations of illuminations, poses, and expressions.

F. Comparison of Average Computing Time

In this section, we compare the processing time of all these methods. To show the computational efficiency, we give the average time (seconds) taken by all the testing samples of every method on the ORL and GT face data sets. Our computing platform is a PC with 2.6 GHz Celeron CPU and 4-GB RAM and the software is MATLAB. Tables VI and VII present the results of the two data sets. These results show that our method has a fast computational speed. This speed is acceptable for real-time applications (such as real-time face recognition application).

G. Analysis of Experimental Results

From the above experiments, the following points can be made. First, our proposed method is suitable for recognition tasks. Especially, under the conditions that face images have

TABLE VI
AVERAGE TIME (S) OF ALL THE TEST SAMPLES ON THE ORL FACE DATA SET

Number of training samples per subject	2	3	4	5	6
Our method	1.90	3.11	4.37	4.58	5.13
CRC	2.44	2.84	2.77	3.16	3.25
L1LS	176.48	233.62	291.62	314.33	312.77
FISTA	12.27	15.37	17.07	18.22	16.03
INNC	3.03	3.58	4.15	4.22	4.10
FCM	2.77	3.36	3.58	3.68	3.35
LDA	2.83	4.14	4.16	3.75	4.05
Homotopy	9.75	9.95	10.41	9.60	8.91
DALM	2069.14	1789.54	1556.63	1323.63	1057.29

TABLE VII
AVERAGE TIME (S) OF ALL THE TEST SAMPLES ON THE GT FACE DATA SET

Number of training samples per subject	3	4	5	6	7	8
Our method	3.94	5.57	9.11	9.33	10.47	12.11
CRC	5.83	6.57	6.96	7.08	7.73	7.93
L1LS	283.03	406.68	446.59	526.33	623.90	627.31
FISTA	20.78	22.08	27.68	31.58	30.95	30.43
INNC	5.90	6.32	7.09	7.75	9.80	11.27
FCM	4.32	6.55	6.32	6.36	6.39	6.84
LDA	1.51	1.31	1.58	1.46	2.18	2.12
Homotopy	15.02	14.68	15.23	16.20	15.97	15.92
DALM	506.82	418.73	403.74	360.90	320.12	295.59

varying illuminations, facial expressions, and poses, our proposed method can still achieve a satisfactory performance. Second, for relatively small-scale data sets (e.g., the ORL data set) and the data sets with slight pose or expression variations (e.g., the Multi-PIE data set), our method is able to achieve a very high recognition accuracy, and the other method also had a good performance. Third, for the parameter in our method, its optimal value varied with the data sets. In addition, the experiments on the Multi-PIE data sets partially showed that in face recognition, the influence of variation of the face appearance, such as expression variation, may be greater than the influence of the number of subjects with frontal poses.

VI. CONCLUSION

With the goal of making the representations of the test sample obtained using different classes as discriminative as possible, we propose a novel sparse representation method, which achieves a noticeable performance in face recognition. Differing from previous studies, this paper designs a simple, computationally efficient, and robust l_2 regularization-based representation algorithm and comprehensively shows its advantages. More importantly, no identical idea was proposed for improving the sparse representation methods

in the previous studies. The proposed method suggests that discriminative presentation can be simply achieved by reducing the correlation of the presentations of the test sample generated from different classes. This shows that in addition to sparsity and collaboration playing important roles in the sparse representation methods, it is also useful to pay attention to decrease the correlation of the presentations of the test sample generated from different classes. The idea presented to enhance the distinctiveness of all the classes may also be useful for improving the l_1 regularization-based sparse representation methods. We will further study this issue in the future. An unsolved problem of the proposed method is that the optimal value of the parameter varies with the data set. In the future, we will address this problem, and attempt to design a procedure to choose the optimal value of the parameter.

ACKNOWLEDGMENT

The authors would like to thank Dr. E. C. Mignot, Shandong University, for the linguistic advice.

REFERENCES

[1] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[3] K. L. Kroeker, "A breakthrough in algorithm design," *Commun. ACM*, vol. 54, no. 9, pp. 13–15, Sep. 2011.

[4] H. Zhang, Y. Zhang, and T. S. Huang, "Pose-robust face recognition via sparse representation," *Pattern Recognit.*, vol. 46, no. 5, pp. 1511–1521, May 2013.

[5] J. Yang, D. Chu, L. Zhang, Y. Xu, and J. Yang, "Sparse representation classifier steered discriminative projection with applications to face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1023–1035, Jul. 2013.

[6] S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Multi-layer group sparse coding—For concurrent image classification and annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 2809–2816.

[7] J. Yang, L. Zhang, Y. Xu, and J.-Y. Yang, "Beyond sparsity: The role of L_1-optimizer in pattern classification," *Pattern Recognit.*, vol. 45, no. 3, pp. 1104–1118, Mar. 2012.

[8] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[9] G. Jing, Y. Shi, D. Kong, W. Ding, and B. Yin, "Image super-resolution based on multi-space sparse representation," *Multimedia Tools Appl.*, vol. 70, no. 2, pp. 741–755, May 2014.

[10] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 457–464.

[11] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 372–386, Feb. 2012.

[12] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[13] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2011, pp. 543–550.

[14] Z. Li, Z. Lai, Y. Xu, J. Yang, and D. Zhang, "A locality-constrained and label embedding dictionary learning algorithm for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2015.2508025.

[15] R. Mazhar and P. D. Gader, "EK-SVD: Optimized dictionary design for sparse representations," in *Proc. IEEE Conf. Pattern Recognit. (ICPR)*, Dec. 2008, pp. 1–4.

[16] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[17] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.

[18] Y. Xu, Q. Zhu, Z. Fan, D. Zhang, J. Mi, and Z. Lai, "Using the idea of the sparse representation to perform coarse-to-fine face recognition," *Inf. Sci.*, vol. 238, no. 20, pp. 138–148, Jun. 2013.

[19] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 471–478.

[20] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, "A two-phase test sample sparse representation method for use with face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 1255–1262, Sep. 2011.

[21] C.-X. Ren, D.-Q. Dai, and H. Yan, "Robust classification using $\ell_{2,1}$-norm based regression model," *Pattern Recognit.*, vol. 45, no. 7, pp. 2708–2718, Jul. 2012.

[22] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1813–1821.

[23] I. Naseem, R. Togneri, and M. Bennamoun, "Robust regression for face recognition," *Pattern Recognit.*, vol. 45, no. 1, pp. 104–118, Jan. 2012.

[24] Y. Xu, X. Zhu, Z. Li, G. Liu, Y. Lu, and H. Liu, "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, Apr. 2013.

[25] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[26] Apr. 2008. *l1_ls: Simple MATLAB Solver for l1-regularized Least Squares Problems*. [Online]. Available: http://web.stanford.edu/~boyd/l1_ls/

[27] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 399–406.

[28] A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Fast ℓ_1-minimization algorithms and an application in robust face recognition: A review," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1849–1852.

[29] A. Y. Yang, S. S. Sastry, A. G. Balasubramanian, S. S. Sastry, and Y. Ma, "Fast ℓ_1-minimization algorithms for robust face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3234–3246, Aug. 2013.

[30] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.

[31] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, pp. 1457–1469, Dec. 2004.

[32] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[33] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3501–3508.

[34] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[35] T. Bai, Y. Li, and X. Zhou, "Monocular human motion tracking with discriminative sparse representation," *Adv. Robot.*, vol. 28, no. 6, pp. 403–414, 2014.

[36] F. Yang, Z. Jiang, and L. S. Davis, "Online discriminative dictionary learning for visual tracking," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2014, pp. 854–861.

[37] S. Zhang, J. Huang, D. Metaxas, W. Wang, and X. Huang, "Discriminative sparse representations for cervigram image segmentation," in *Proc. IEEE Int. Symp. Biomed. Image., Nano Macro*, Apr. 2010, pp. 133–136.

[38] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[39] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. 2nd IEEE Workshop Appl. Comput. Vis.*, Dec. 1994, pp. 138–142.

[40] N. Goel, G. Bebis, and A. Nefian, "Face recognition experiments with random projection," *Proc. SPIE*, vol. 5779, pp. 426–437, Apr. 2005.

[41] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable

lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
[42] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, May 2015.
[43] Y. Xu, Q. Zhu, Y. Chen, and J.-S. Pan, "An improvement to the nearest neighbor classifier and face recognition experiments," *Int. J. Innov. Comput., Inf. Control*, vol. 9, no. 2, pp. 543–554, Feb. 2013.
[44] Z. Liu, J. Pu, T. Huang, and Y. Qiu, "A novel classification method for palmprint recognition based on reconstruction error and normalized distance," *Appl. Intell.*, vol. 39, no. 2, pp. 307–314, Feb. 2013.
[45] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[46] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, Sep. 2008, pp. 1–8.

Yong Xu (M'06–SM'15) received the B.S. and M.S. degrees from the PLA University of Science and Technology, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology, Nanjing, China, in 2005.

He is currently a Professor with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China. His current research interests include pattern recognition, biometrics, bioinformatics, machine learning, image processing, and video analysis.

Zuofeng Zhong received the B.S. degree in computer science and technology from Guangxi Normal University, Guilin, China, in 2007, and the M.S. degree in computer science and technology from Guangdong University of Technology, Guangzhou, China, in 2011. He is currently pursuing the Ph.D. degree in computer science and technology with the Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China.

His current research interests include pattern recognition, image processing, and video analysis.

Jian Yang (M'06) received the B.S. degree in mathematics from Xuzhou Normal University, Xuzhou, China, in 1995, the M.S. degree in applied mathematics from Changsha Railway University, Changsha, China, in 1998, and the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2002.

He was a Post-Doctoral Researcher with the University of Zaragoza, Zaragoza, Spain, in 2003. From 2004 to 2006, he was a Post-Doctoral Fellow with the Biometrics Center, The Hong Kong Polytechnic University, Hong Kong. From 2006 to 2007, he was a Post-Doctoral Fellow with the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, USA. He is currently a Professor with the School of Computer Science and Technology, NUST. He has authored more than 50 scientific papers on pattern recognition and computer vision. His journal papers have been cited more than 1200 times on the ISI Web of Science, and 2000 times on Google Scholar. His current research interests include pattern recognition, computer vision, and machine learning.

Jane You received the B.Eng. degree in electronics engineering from Xi'an Jiaotong University, Xi'an, China, in 1986, and the Ph.D. degree in computer science from La Trobe University, Melbourne, VIC, Australia, in 1992.

She was a Lecturer with the University of South Australia, Adelaide SA, Australia, and a Senior Lecturer with Griffith University, Nathan, QLD, Australia, from 1993 to 2002. She is currently a Full Professor with The Hong Kong Polytechnic University, Hong Kong. Her current research interests include image processing, pattern recognition, medical imaging, biometrics computing, multimedia systems, and data mining

Dr. You received the French Foreign Ministry International Fellowship Award from 1993 to 1994.

David Zhang (F'09) received the B.S. degree in computer science from Peking University, Beijing, China, in 1974, the M.S. degree in computer science and the Ph.D. degree from the Harbin Institute of Technology (HIT), Harbin, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 1994.

He was a Post-Doctoral Fellow with Tsinghua University, Beijing, from 1986 to 1988, and then an Associate Professor with Academia Sinica, Beijing. He is currently the Chair Professor with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, where he is the Founding Director of the Biometrics Technology Center supported by the Hong Kong Government in 1998. He serves as a Visiting Chair Professor with Tsinghua University, and an Adjunct Professor with Shanghai Jiao Tong University, Shanghai, China, Beihang University, Beijing, HIT, and the University of Waterloo. He has authored more than ten books and 160 journal papers.

Dr. Zhang is a Croucher Senior Research Fellow, a Distinguished Speaker of the IEEE Computer Society, and a fellow of the International Association of Pattern Recognition. He is the Founder and Editor-in-Chief of the *International Journal of Image and Graphics*, a Book Editor of the *Springer International Series on Biometrics*, an Associate Editor of more than ten international journals, including the IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS A/SYSTEMS MAN AND CYBERNETICS-C/PATTERN RECOGNITION, and the Chair of the IEEE/CIS Technical Committee on Intelligent Systems Application. He was an Organizer of the International Conference on Biometrics in 2004 and 2006.