IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 4, 2018, accepted November 15, 2018, date of publication November 27, 2018, date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2883527

A New Discriminative Collaborative Neighbor Representation Method for Robust Face Recognition

JIANPING GOU[1], (Member, IEEE), LEI WANG[1], ZHANG YI[2], (Fellow, IEEE), JIANCHENG LV[2], QIRONG MAO[1], AND YUN-HAO YUAN[3]

[1]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China
[2]College of Computer Science, Sichuan University, Chengdu 610065, China
[3]College of Information Engineering, Yangzhou University, Yangzhou 225127, China

Corresponding author: Jianping Gou (goujianping@ujs.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61502208, in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20150522, in part by the International Postdoctoral Exchange Fellowship Program from the Chinese Postdoctoral Council under Grant 20180051, in part by the China Postdoctoral Science Foundation under Grant 2015M570411, in part by the Research Foundation for Talented Scholars of Jiangsu University under Grant 14JDG037, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX18_2259, and in part by the Open Foundation of Artificial Intelligence Key Laboratory of Sichuan Province under Grant 2017RYJ04.

ABSTRACT As the representative one of representation-based classification (RBC) methods, collaborative RBC (CRC) has drawn much attention in pattern recognition and machine learning recently. Moreover, the collaborative representation-based face recognition has been extensively studied because of the effective classification performance of CRC. CRC collaboratively represents each query sample as the linear combination of all the training samples and then classifies the query sample according to the categorical representation-based distances. However, most variants of CRC cannot fully consider the locality and discrimination of data and cannot well handle the noise data, which has negative effect on real-world classification problems, such as face recognition. In this paper, a new discriminative collaborative neighbor representation (DCNR) method for face recognition is proposed by integrating class discrimination and data locality. In the proposed method, the locality of data constrains collaborative representation of each query sample to find representative nearest samples of the query sample. Moreover, the class discrimination regularization is taken into account by employing the representation of each class for each query sample. Due to the existing noises, such as corruptions and occlusions in face recognition, we further propose robust DCNR (R-DCNR) for robust classification by using the ℓ_1-norm representation fidelity. Extensive experiments on face databases demonstrate that the proposed methods achieve competitive classification performance, compared to the state-of-the-art representation-based classification methods.

INDEX TERMS Representation-based classification, collaborative representation, sparse representation, face recognition.

I. INTRODUCTION

Representation-based classification (RBC) in representation learning has been extensively studied in the field of pattern recognition and machine learning. Recently, there are many existing representation-based classification methods that are used in face recognition [1]–[7] and image classification [8]–[11]. Generally speaking, a great many RBC methods are mainly divided into two representative types: sparse representation-based classification (SRC) [1] and collaborative representation-based classification (CRC) [2]. Moreover, some new overviews about the progresses and the applications of representation learning have been presented in [12]–[16].

In pattern recognition, sparse representation-based classification (SRC) was first introduced in [1] for face recognition and then has significantly attracted much attention. Since sparse representation holds good discrimination of data [1], [17], many variants of SRC have been proposed for

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

good classification nowadays, such as in [6], [7], [18]–[21], and [24]–[27]. To improve the classification performance of benchmark SRC, some weighted SRC methods were proposed by considering the localities of data as the weights that constrained sparse representation coefficients [18]–[20]. Moreover, using the locality constraints on sparse representation coefficients, locality-sensitive dictionary learning for SRC was proposed in [21]–[23] for enhancing the power of discrimination. In [24], sparsity and correlation of data were simultaneously considered to design adaptive sparse representation for classification. In [25], fisher discrimination dictionary learning on the basis of the fisher discrimination criterion was proposed by jointly employing the discrimination of both representation residuals and representation coefficients. Due to the discrimination of representation coefficients, the extensions of SRC were introduced by adopting the representation coefficients to design the classification decision [26], [27]. Thus, sparse representation-based classification has been one of the representative methods in pattern recognition and especially performs well in face recognition.

Different from SRC that uses the l_1 regularization of representation coefficients, the collaborative representation-based classification (CRC) as another representative RBC method was proposed by generally employing l_2 regularization of representation coefficients [2]. CRC results in a closed-form solution which is efficient for pattern classification. As argued in [2], l_2-norm collaboration rather than l_1-norm sparsity makes SRC has the promising classification performance. Furthermore, sparse representation is regarded as one of collaborative representation in [2] and [4]. Recently, there are many extensions of CRC [3]–[6], [8], [10], [28]–[36]. In [4], the discriminant nature of CRC was argued. To overcome outliers and noises, the robust CRC [28] was proposed for dealing with occlusions and corruptions in face recognition. Considering that the intrinsic mechanism of CRC was unclear, an explanation from the perspective of probability was given and the probabilistic collaborative representation-based classifier (ProCRC) which maximizes the likelihood that the query sample belongs to each class was proposed in [5]. It has been proven that ProCRC has superior classification performance over CRC. To enhance the classification performance of ProCRC, the robust ProCRC (R-ProCRC) was also proposed by using l_1-norm fidelity term. In [29], the discriminative CRC based on dictionary learning was proposed to strengthen the discrimination of data. Using the idea of coarse to fine, several two-phase CRC methods were developed in [3], [6], and [30]–[32]. Similar to two-phase CRC, hierarchical collaborative representation using a two-stage classifier was proposed for classification in [35]. To emphasize the discrimination among classes, the competitive CRC was proposed lately in [33] and [34]. In [36], the integrating conventional and inverse collaborative representation was proposed for face recognition. According to the above literatures about CRC, it has been shown that CRC achieves good classification performance, especially in face recognition.

In fact, enhancing the discrimination of data and considering the locality of data have an effective impact on improving the performance of CRC. To fully use the discrimination of data, a new discriminative sparse representation method (DSRC) with l_2 regularization of collaborative representation was proposed by using the discrimination of the class-specific representations [37]. DSRC not only obtains an efficient closed-form solution, but also has sparsity nature of l_1 regularization of the sparse representation coefficients. To overcome the noises in face recognition, the antinoise sparse representation method as an extended variant of DSRC was proposed by using l_1-norm based representation fidelity term [38]. To well utilize the locality of data, some variants of CRC were proposed in [39]–[43]. In the weighted CRC [39], the locality of data is reflected by considering the local similarity distances between training samples and each testing sample. Just like the weighted CRC in [39], the collaborative neighbor representation-based classification (CNRC) [40] was proposed by the locality constraints and the regularization of the representation coefficients. In [41], CNRC was extended to the two phase collaborative neighbor representation method. In [42], the supervised neighborhood regularized collaborative representation method was proposed by considering the class-specific neighborhood structures of data. In [43], the proposed locality preserving collaborative representation method is also an extension of CNRC.

Inspired by the ideas of both DSRC and CNRC, we propose a new discriminative collaborative neighbor representation method (DCNR) for face recognition. Since the locality of the face data with variations in poses, expressions and illuminations is very important for classification [41]–[43], the proposed DCNR integrates the locality of data and the discrimination information of the class-specific representations. The locality of data in DCNR is reflected by the local similarity distances between all training samples and each testing sample and then constrains the representation coefficients. The discrimination information in DCNR is mainly enhanced by minimizing the class-specific representation and degrading the correlation among classes. Since noises such as occlusions and corruptions are often contained in face data, we extend the proposed DCNR to the robust DCNR (R-DCNR) for robust face recognition. In the R-DCNR, we adopt ℓ_1-norm of coding residual to characterize the representation fidelity to enhance the discrimination from noise face data [5], [38]. To extensively verify the effectiveness of the proposed methods, experiments on seven face image databases are conducted. In comparisons with the state-of-the-art RBC methods, the experimental results show that the proposed methods achieve the competitive classification performance.

The remainder of this article is organized as follows. Section II presents the related methods. Section III introduces the proposed methods in details. Section IV gives the detailed analyses of the proposed methods. Section V reports the experimental results. Finally, the article is concluded in Section VI.

II. RELATED METHODS

This section briefly outlines two related RBC methods as background knowledge of the proposed methods. For convenience, some notations are first introduced as follows. Let a matrix $X = [X_1, X_2, \ldots, X_c, \ldots, X_L] \in \mathbb{R}^{m\times N}$ denote a set of all N training samples from L classes $\{1, 2, \ldots, L\}$. $X_c = [x_{n(c-1)+1}, x_{n(c-1)+2}, \ldots, x_{nc}]$ is the subset of training samples from the cth class, where n is the number of each class-specific training samples and $N = nL$. In the representation-based classification methods, the given testing sample $y \in \mathbb{R}^m$ can be approximatively linearly represented as

$$y \approx XS, \tag{1}$$

where $S = [s_1, s_2, \ldots, s_N]^T$ is the representation coefficient vector.

A. CNRC

The collaborative neighbor representation-based classification (CNRC) [40] as an efficient RBC method considers the locality of data through the assumption of locally linear embedding (LLE) [44]. CNRC employs the local similarity distances between each testing sample and all the training samples to reflect the locality of data and to constrain the representation coefficients. It can find the representative nearest training samples of each testing sample to strengthen the discrimination power. The model of CNRC is formulated as

$$S = \arg\min_S \left\{ \frac{1}{2}\Big(\| y - XS \|_2^2 + \gamma \sum_{i=1}^{N} s_i^2 \cdot \| y - x_i \|_2^2 + \lambda \| S \|_2^2 \Big) \right\}, \tag{2}$$

where γ and λ are small positive regularization parameters and s_i denotes the representation coefficient corresponding to the ith training sample x_i. The local similarity distance $d_i = \| y - x_i \|_2^2$ between x_i and y constrains the representation coefficient s_i of x_i. The closer d_i could lead to a larger representation coefficient s_i. That is, a farther training sample makes less contribution to reconstructing y.

After the representation coefficient vector S is obtained, the given testing sample y can be classified into the class which has minimum reconstruct error $r_c(y) = \| y - X_c S_c \|_2^2 / \| S_c \|_2^2$, where $S_c \in \mathbb{R}^n$ denotes the representation coefficients from the cth class and $\| S_c \|_2^2$ can increase the discrimination for classification [28].

B. DSRC

The discriminative sparse representation method (DSRC) with l_2 regularization of representation is a very efficient, easily solvable and robust method for face recognition [37]. Moreover, compared to l_2 regularization-based representation methods, DSRC holds the sparsity property which is useful for choosing the representative training samples to represent the testing sample. The model of DSRC is defined as

$$S = \arg\min_S \left\{ \| y - XS \|_2^2 + \beta \sum_{i=1}^{L}\sum_{j=1}^{L} \| X_i S_i + X_j S_j \|_2^2 \right\}, \tag{3}$$

where β is a positive constant, $S_i = [s_{n(i-1)+1}, s_{n(i-1)+2}, \ldots, s_{in}]$, and $X_i S_i$ denotes the class-specific representation, $i \in \{1, 2, \ldots, L\}$.

In Eq. (3), the second term as a regularization term can make class-specific representations from different classes have the low correlations and can own sparsity to enhance the discrimination power. In DSRC, the given testing sample y is classified into the class with the minimum reconstruct error $r_c(y) = \| y - X_c S_c \|_2^2$.

III. THE PROPOSED METHODS

In this section, we describe the proposed methods in details. Inspired by the ideas of CNRC [40] and DSRC [37], we simultaneously consider the discrimination from class-specific representations and the locality of data in the CRC model and propose the discriminative collaborative neighbor representation method (DCNR). The goal of DCNR is to discriminatively reconstruct the testing sample y by learning the representative nearest training samples from the same class which y belongs to and to correctly classify y. To strengthen the robustness of the classification performance in the case of noises, such as in face recognition with partial corruptions and occlusions, we also extend the DCNR to the robust one (R-DCNR) that utilizes the ℓ_1-norm of coding residual to characterize the representation fidelity.

A. THE DCNR METHOD

For a given testing sample y, the model of DCNR is designed as follows:

$$S = \arg\min_S \left\{ \| y - XS \|_2^2 + \beta \sum_{i=1}^{L}\sum_{j=1}^{L} \| X_i S_i + X_j S_j \|_2^2 + \gamma \sum_{h=1}^{N} s_h^2 \cdot \| y - x_h \|_2^2 + \lambda \| S \|_2^2 \right\}, \tag{4}$$

where s_h denotes the hth element in representation coefficient vector S, x_h is the hth training sample and β, γ, λ are small positive regularization parameters for balancing the influence between the regularization terms and the loss function. The second term in Eq. (4) is the discrimination constraint and the third term is the locality constraint. In Eq. (4), $S = [S_1, S_2, \cdots, S_L]^T$ and $S_k \in \mathbb{R}^n$ denotes the representation coefficients of training samples from the kth class. When they are set to be the suitable values, Eq. (4) can obtain an optimized representation coefficient vector S.

Since the proposed objective function of DCNR is convex and differentiable, its solution can be easily obtained by the derivation. Firstly, the benchmark CRC model $f_1 = \| y - XS \|_2^2 + \lambda \| S \|_2^2$ in Eq. (4) can collaboratively reconstruct the testing sample y with essential discrimination information for classification [2]. The derivation of f_1 with respect to the collaborative representation coefficient vector S is computed as:

$$\begin{aligned}\frac{df_1}{dS} &= \frac{d\left(\parallel y - XS \parallel_2^2 + \lambda \parallel S \parallel_2^2\right)}{dS} \\ &= 2\left((X^TX+\lambda I)S - X^Ty\right),\end{aligned} \tag{5}$$

where I denotes an identity matrix. Through the benchmark CRC model, each training sample has the corresponding contribution to reconstructing the testing sample y. After acquiring the solution from Eq. (5), the training samples from the true class that y belongs to could have larger representation coefficients in general.

Secondly, the main goal of $f_2 = \beta \sum_{i=1}^{L}\sum_{j=1}^{L} \parallel X_iS_i + X_jS_j \parallel_2^2$ in Eq. (4) is to enhance the discrimination among different classes by using the class-specific representations. Since $S = [S_1, S_2, \cdots, S_L]^T$, we should compute the derivative df_2/dS through all partial derivatives $\partial f_2/\partial S_k$ $(k = 1, 2, \ldots, L)$. That is

$$\frac{df_2}{dS} = \left[\frac{\partial f_2}{\partial S_1}, \frac{\partial f_2}{\partial S_2}, \ldots, \frac{\partial f_2}{\partial S_L}\right]^T. \tag{6}$$

To compute $\partial f_2/\partial S_k, f_2$ can be first rewritten as follows:

$$\begin{aligned}f_2 &= \beta\left(\sum_{i=1}^{L}\sum_{j=1}^{L} \parallel X_iS_i + X_jS_j \parallel_2^2\right) \\ &= \beta\left(\sum_{i=1}^{L}\sum_{j\neq k}^{L} \parallel X_iS_i + X_jS_j \parallel_2^2 + \sum_{i=1}^{L} \parallel X_iS_i + X_kS_k \parallel_2^2\right) \\ &= \beta\left(\sum_{i\neq k}^{L}\sum_{j\neq k}^{L} \parallel X_iS_i + X_jS_j \parallel_2^2 + \sum_{j\neq k}^{L} \parallel X_kS_k + X_jS_j \parallel_2^2\right. \\ &\quad \left. + \sum_{i\neq k}^{L} \parallel X_iS_i + X_kS_k \parallel_2^2 + \parallel 2X_kS_k \parallel_2^2\right) \\ &= \beta\left(\sum_{i\neq k}^{L}\sum_{j\neq k}^{L} \parallel X_iS_i + X_jS_j \parallel_2^2 + 2\sum_{i\neq k}^{L} \parallel X_iS_i + X_kS_k \parallel_2^2\right. \\ &\quad \left. + 4 \parallel X_kS_k \parallel_2^2\right).\end{aligned} \tag{7}$$

Using Eq. (7), we can easily obtain the partial derivative $\partial f_2/\partial S_k$ as

$$\begin{aligned}\frac{\partial f_2}{\partial S_k} &= \beta\frac{\partial}{\partial S_k}\left(\sum_{i\neq k}^{L}\sum_{j\neq k}^{L} \parallel X_iS_i + X_jS_j \parallel_2^2 + 4 \parallel X_kS_k \parallel_2^2\right. \\ &\quad \left. + 2\sum_{i\neq k}^{L} \parallel X_iS_i + X_kS_k \parallel_2^2\right) \\ &= \beta\left(8X_k^TX_kS_k + 4\sum_{i\neq k}^{L} X_k^T(X_iS_i + X_kS_k)\right) \\ &= 4\beta\left(2X_k^TX_kS_k + X_k^T\sum_{i\neq k}^{L} X_iS_i + (L-1)X_k^TX_kS_k\right) \\ &= 4\beta\left(LX_k^TX_kS_k + X_k^T\sum_{i=1}^{L} X_iS_i\right) \\ &= 4\beta X_k^T(LX_kS_k + XS).\end{aligned} \tag{8}$$

Thus, df_2/dS in Eq. (6) can be easily achieved as

$$\begin{aligned}\frac{df_2}{dS} &= \begin{bmatrix} 4\beta X_1^T(LX_1S_1 + XS) \\ \vdots \\ 4\beta X_L^T(LX_LS_L + XS)\end{bmatrix} \\ &= 4\beta L\begin{bmatrix} X_1^TX_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & X_L^TX_L\end{bmatrix} S + 4\beta X^TXS \\ &= 4\beta LM + 4\beta X^TXS \\ &= 4\beta(LM + X^TX)S,\end{aligned} \tag{9}$$

where

$$M = \begin{bmatrix} X_1^TX_1 & & 0 \\ & \ddots & \\ 0 & & X_L^TX_L\end{bmatrix}, \tag{10}$$

is a block diagonal matrix and $O \in \mathbb{R}^{n\times n}$ is a zero matrix.

Thirdly, let $f_3 = \gamma \sum_{h=1}^{N} s_h^2 \cdot \parallel y - x_h \parallel_2^2)$ in Eq. (4). It is the locality regularization term that can make collaborative representation to seek the representative nearest training samples of the testing samples y. To solve the representation coefficient vector S, we rewrite f_3 as

$$\begin{aligned}f_3 &= \gamma \left\|\begin{bmatrix} \parallel y - x_1 \parallel_2 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \ldots & \parallel y - x_N \parallel_2\end{bmatrix}\begin{bmatrix} s_1 \\ \vdots \\ s_N\end{bmatrix}\right\|_2^2 \\ &= \gamma \parallel DS \parallel_2^2,\end{aligned} \tag{11}$$

where

$$D = \begin{bmatrix} \parallel y - x_1 \parallel_2 & & 0 \\ & \ddots & \\ 0 & & \parallel y - x_N \parallel_2\end{bmatrix}. \tag{12}$$

$D \in \mathbb{R}^{N\times N}$ is the locality weight matrix of S that can facilitate the closer training samples of y to have more contributions to reconstructing y. Using Eq. (11), df_3/dS is easily computed as

$$\begin{aligned}\frac{df_3}{dS} &= \frac{d \parallel DS \parallel_2^2}{dS} \\ &= 2D^TDS.\end{aligned} \tag{13}$$

Finally, using Eqs. (5), (9) and (13), the optimal solution of Eq. (4) can be obtained by the following function

$$\left((X^TX + \lambda I)S - X^Ty\right) + 2\beta(LM + X^TX)S + D^TDS = 0. \tag{14}$$

Obviously, the representation coefficient vector S in Eq. (4) is optimally solved as

$$S = (X^T X+\gamma D^T D+2\beta(X^T X+LM)+\lambda I)^{-1}X^T y. \tag{15}$$

Using the optimal representation coefficient vector S, the testing sample can be discriminatively reconstructed by the representative nearest training samples among all the classes. Then, the testing sample y is classified into the class that has minimum reconstruction residual as

$$l = \arg\min_c \ \| y - X_c S_c \|_2^2, \tag{16}$$

where $c \in \{1, 2, \dots, L\}$. The procedure of the proposed DCNR method is briefly presented in Algorithm 1.

Algorithm 1 Discriminative Collaborative Neighbor Representation Method (DCNR)

Input: A training sample set $X = [X_1, X_2, \dots, X_L] \in \mathbb{R}^{m\times N}$, a testing sample $y \in \mathbb{R}^m$ and the positive constants λ, β, γ.

Output: Representation coefficient vector S and the class label l of y.

Step 1: Normalize each column of X.

Step 2: Calculate representation coefficient vector S

$$S = (X^T X + \gamma D^T D + 2\beta(X^T X + LM) + \lambda I)^{-1}X^T y.$$

Step 3: Obtain class-specific reconstruction residual

$$d(c) = \| y - X_c S_c \|_2^2,$$

where $c \in \{1, 2, \dots, L\}$.

Step 4: Assign y into the lth class by

$$l = \arg\min_c d(c).$$

B. THE ROBUST DCNR METHOD

Although many CRC methods have good performance for image classification, they often have less robustness in the case of noises. For example, in face recognition, the classification performance of CRC is always degraded by the partial corruptions and occlusions of face images. As argued in [5] and [38], the representation fidelity term that is constrained with l_1-norm regularization can achieve the robust classification performance in face recognition with the corruption and occlusion noises. To enhance the robustness of the classification performance of the proposed DCNR, we introduce the robust variant of DCNR (R-DCNR) by enforcing the l_1-norm coding residual on the representation fidelity. Then, the model of R-DCNR is defined as

$$S = \arg\min_S \Bigg\{ \| y - XS \|_1 + \beta \sum_{i=1}^{L}\sum_{j=1}^{L} \| X_i S_i + X_j S_j \|_2^2 + \gamma \sum_{h=1}^{N} s_h^2 \cdot \| y - x_h \|_2^2 + \lambda \| S \|_2^2 \Bigg\}. \tag{17}$$

Obviously, compared to DCNR, l_1-norm instead of l_2-norm is used for the representation fidelity term in R-DCNR.

Since the R-DCNR model cannot directly obtain closed-form solution, we employ iterative reweighted least square (IRLS) [45] technique to reformulate this model in order to easily compute the solution. In IRLS, a diagonal weight matrix is first defined as

$$W_X = \begin{bmatrix} 1/|X(1,:)S - y_1| & & 0 \\ & \ddots & \\ 0 & & 1/|X(m,:)S - y_m| \end{bmatrix}, \tag{18}$$

where $X(i, :)$ denotes the ith row of training sample matrix X and y_i is the ith element in the testing sample vector y. Then, the objective function in Eq. (17) is rewritten as

$$S = \arg\min_S \Bigg\{ (y - XS)^T W_X (y-XS) + \gamma \sum_{h=1}^{N} s_h^2 \cdot \| y - x_h \|_2^2 + \beta \sum_{i=1}^{L}\sum_{j=1}^{L} \| X_i S_i + X_j S_j \|_2^2 + \lambda \| S \|_2^2 \Bigg\}. \tag{19}$$

Through Eq. (19), the representation coefficients S in Eq. (17) can be easily solved as

$$S = (X^T W_X X+\gamma D^T D+2\beta(X^T X + LM) + \lambda I)^{-1}X^T W_X y. \tag{20}$$

The weight matrix W_X and the representation coefficient vector S are iteratively updated until the preseted iterations or convergence are reached. In Algorithm 2, we summarize the proposed R-DCNR method.

Algorithm 2 Robust Discriminative Collaborative Neighbor Representation Method (R-DCNR)

Input: A training sample set $X = [X_1, X_2, \dots, X_L] \in \mathbb{R}^{m\times N}$, a testing sample $y \in \mathbb{R}^m$ and the positive constants λ, β, γ.

Output: Representation coefficient vector S and the class label l of y.

Step 1: Perform steps 1 and 2 in Algorithm 1 to determine the solution S_0 by DCNR.

Step 2: Initialize S with S_0.

Step 3: Iteratively solve weight matrix W_X and representation coefficient vector S

while

$$\| W_X^{(t+1)} - W_X^{(t)} \|_2^2 \leq \delta_W,$$

or

$$t \leqslant T - 1,$$

do

update weight matrix W_X by Eq. (18).

update representation coefficient vector S by Eq. (20).

Note that $W_X^{(t)}$ is the weight matrix in the tth iteration, δ_W is the threshold value and T is the preseted number of iterations.

end while

Step 4: Classify y with steps 3 and 4 in Algorithm 1.

IV. THE ANALYSES OF THE PROPOSED METHODS

In this section, the basic rationale and the main computational complexities of the proposed methods are analyzed in details.

A. RATIONALE OF THE PROPOSED METHODS

The basic idea of the proposed DCNR is mainly based on both CNRC [40] and DSRC [37]. In the model of DCNR, the locality property of CNRC and the discrimination property from the class-specific representations of DSRC are jointly integrated for classification. For clearly emphasizing the rationale of the proposed DCNR method, we analyze it from the perspective of the locality regularization term and the discrimination regularization term.

It has been argued that the locality of data plays an important role in CRC [39]–[43], but most variants of CRC cannot consider it. In the proposed DCNR, the locality of data is considered by assuming that a testing sample can be truly reconstructed by the nearest training samples, most of which come from the same class y belongs to. It is reflected by the term $\sum_{h=1}^{N} s_h^2 \cdot \parallel y - x_h \parallel_2^2$ in the proposed model. That is to say, each local similarity distance between the testing sample y and the training sample x_h is taken into account and constrains the representation coefficient of x_h. Note that this locality term is also used in CNRC. Minimizing $\sum_{h=1}^{N} s_h^2 \cdot \parallel y - x_h \parallel_2^2$ can make the representation coefficient s_h corresponding to the training sample x_h very large when the local similarity distance between x_h and y is very small. Thus, using the locality term, the proposed DCNR method can find the nearest training samples to represent the testing sample y with large coefficients. The locality property implies that nearest training samples can make more contributions to representing y and then the discrimination ability of DCNR is enhanced.

To strengthen the power of pattern discrimination, the discrimination regularization term $\parallel X_iS_i + X_jS_j \parallel_2^2$ that is used in DSRC is adopted in the proposed DCNR method. In order to analyze the property of this term, we rewrite it as $\parallel X_iS_i \parallel_2^2 + \parallel X_jS_j \parallel_2^2 + 2(X_iS_i)^T(X_jS_j)$. Then, to minimize $\parallel X_iS_i + X_jS_j \parallel_2^2$ is to simultaneously minimize $\parallel X_iS_i \parallel_2^2$ and $(X_iS_i)^T(X_jS_j)$. Firstly, minimizing $\parallel X_iS_i \parallel_2^2$ means that each class-specific representation has a small norm in order to obtain the categorical discrimination. Furthermore, to minimize $(X_iS_i)^T(X_jS_j)$ in fact is to degrade the correlations among classes. The low correlations among classes can improve the discrimination among different classes. In addition, in Eq. (9) the matrix $M = diag(X_1^TX_1, \ldots, X_k^TX_k, \ldots, X_L^TX_L)$ obtained from $\parallel X_iS_i + X_jS_j \parallel_2^2$ can enhance the correlations between the samples within the same class for good classification. Thus, the discrimination regularization term $\parallel X_iS_i + X_jS_j \parallel_2^2$ can well improve classification performance of the proposed DCNR method.

In the proposed DCNR, we well integrate both the locality term and discrimination term for good classification. The superior properties of both terms can be well intuitively presented by comparing DCNR to CRC and by comparing DCNR to CNRC to verify effectiveness of the discrimination regularization, and by comparing DCNR to DSRC to verify effectiveness of the locality regularization. The pair-wise comparisons have been done from the point view of the representation coefficient of each training sample, the sums of the class-specific representation coefficients and the class-specific residuals. The examples of the comparisons of DCNR and CRC, DCNR and CNRC, DCNR and DSRC are shown in Fig. 1, 2 and 3, respectively. Note that the experimental examples have been conducted on the AR face database shown in V-A. In three figures, the representation coefficient of each training sample, the sums of the class-specific representation coefficients and the class-specific residuals from the true class of the given testing samples are indicated in red. Meantime, the representation coefficient of each training sample, the sums of the class-specific representation coefficients and the class-specific residuals from the class that the methods wrongly classify the given testing sample into are indicated in green.

The example of the comparison of DCNR and CRC is shown in Fig. 1. Since the classification decision rules of DCNR and CRC are the reconstructive residuals, it is clear that DCNR correctly classifies the given testing sample and CRC wrongly classifies the testing sample in Fig. 1. From the representation coefficients of all training samples in Figs. 1(a) and 1(b), we can see that the variants of CRC can collaboratively reconstruct the given testing sample. Although the sum of collaborative representation coefficients from each class in Figs. 1(c) and 1(d) can well discriminate the different classes, DCNR has the largest sum of coefficients from the true class of the given testing sample, but CRC has the largest sum of coefficients from its different class. In Figs. 1(e) and 1(f), the given testing sample is correctly determined by the minimum residual among all the classes in DCNR but mistakenly in CRC. From the sums of the class-specific coefficients and the class-specific residuals, we can see that the largest the sum of categorical coefficients correspond to the minimum residual. Thus, the experimental comparison in Fig. 1 means that the locality and discrimination constraints considered in the proposed method are very useful for classification.

The comparison of the proposed DCNR and CNRC in Fig. 2 is to illustrate the availability of the discrimination term $\parallel X_iS_i + X_jS_j \parallel_2^2$, when both DCNR and CNRC have the same locality term $\sum_{h=1}^{N} s_h^2 \cdot \parallel y - x_h \parallel_2^2$. We can see that DCNR and CNRC as two variants of CRC collaboratively represent the given testing sample. Since DCNR and CNRC use the class-specific residuals as the classification decision, it is obvious that the proposed DCNR rightly classifies the given testing sample into the true class with the largest sum of coefficients and the minimum residual among all classes, but CNRC mistakenly determine the class of the testing sample. This experimental fact in Fig. 2 could indicate that DCNR with the discrimination term has more power of pattern discrimination than CNRC.

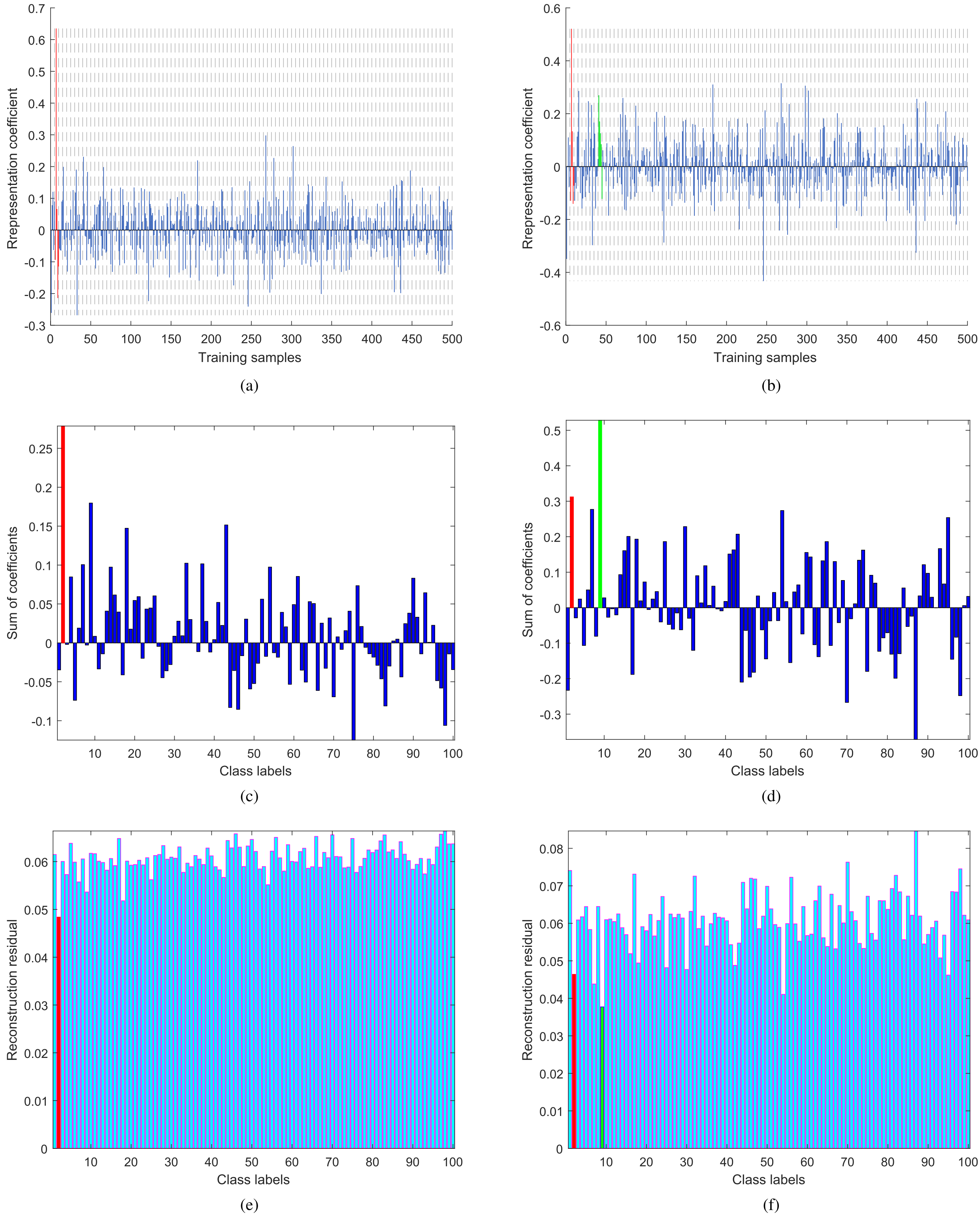


FIGURE 1. The representation coefficient of each training sample (Fig1. (a) and (b)), the sums of the class-specific representation coefficients (Fig1. (c) and (d)), and the class-specific residuals (Fig1. (e) and (f)) for one given testing sample from class 2 by using DCNR and CRC on AR. (a) DCNR. (b) CRC. (c) DCNR. (d) CRC. (e) DCNR. (f) CRC.

To given the intuitive availability of the locality term $\sum_{h=1}^{N} s_h^2 \cdot \parallel y - x_h \parallel_2^2$, the comparison of DCNR and DSRC is displayed in Fig. 3 when both DCNR and DSRC have the same discrimination term $\parallel X_i S_i + X_j S_j \parallel_2^2$. Clearly, these two variants of CRC collaboratively represent the given testing sample. We can observe that DCNR discriminatively and correctly classifies the testing sample into the true class

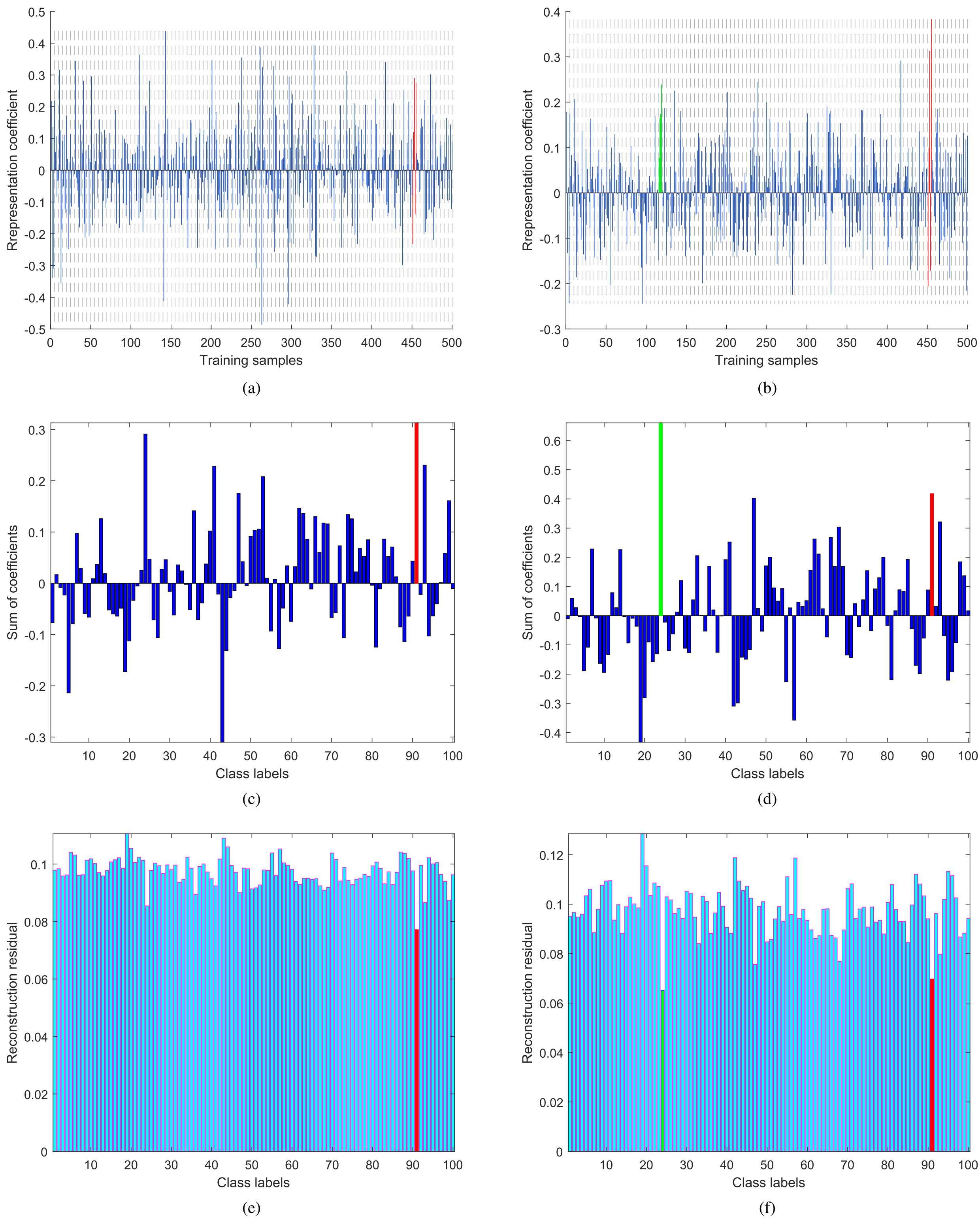


FIGURE 2. The representation coefficient of each training sample (Fig2. (a) and (b)), the sums of the class-specific representation coefficients (Fig2. (c) and (d)), and the class-specific residuals (Fig2. (e) and (f)) for one given testing sample from class 91 by using DCNR and CNRC on AR. (a) DCNR. (b) CNRC. (c) DCNR. (d) CNRC. (e) DCNR. (f) CNRC.

with the largest sum of coefficients and the minimum residual among all the classes, but DSRC wrongly classifies the testing sample. This means the locality term can make the testing sample find the representative nearest training samples so as to improve the discrimination ability of the proposed DCNR.

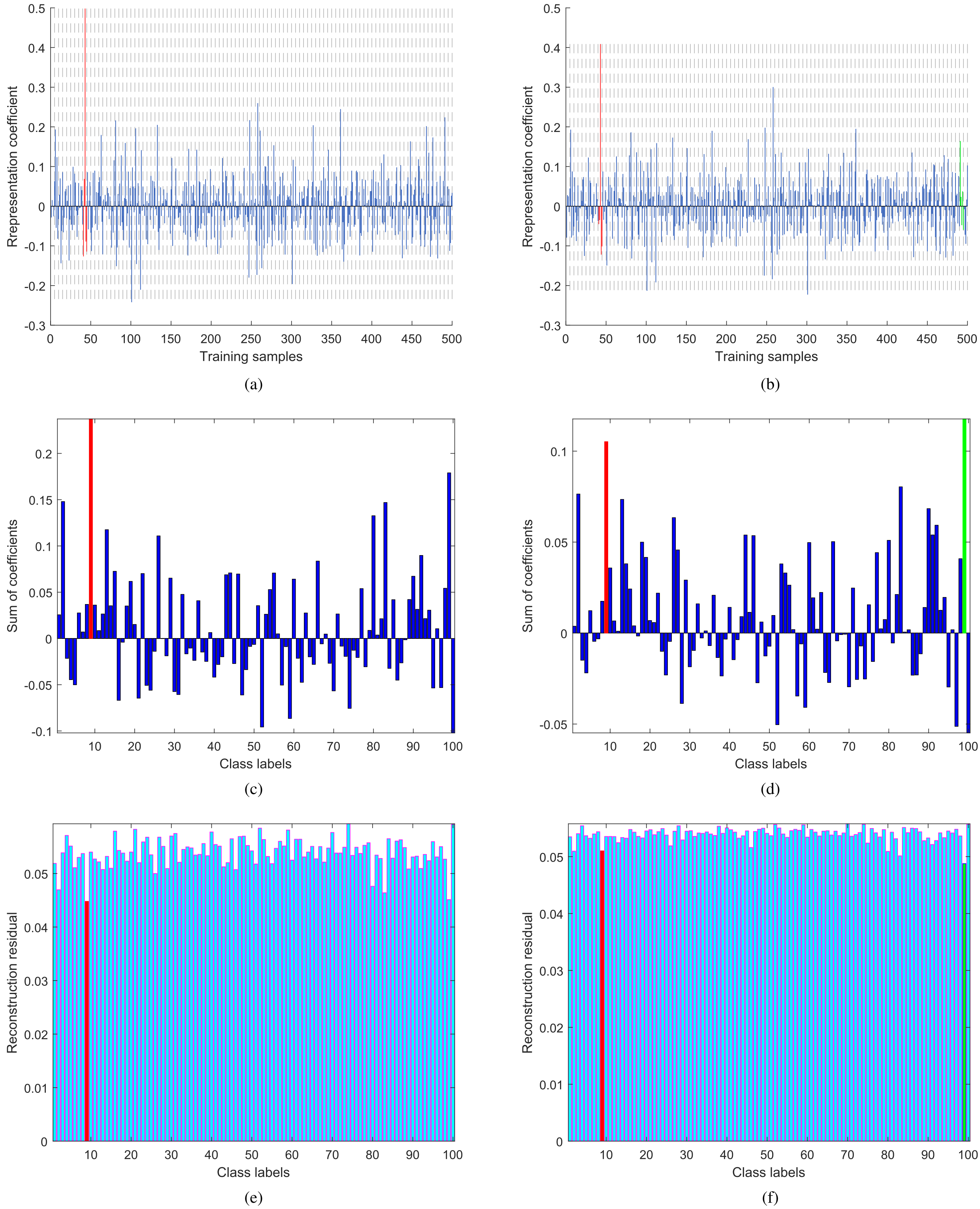


FIGURE 3. The representation coefficient of each training sample (Fig3. (a) and (b)), the sums of the class-specific representation coefficients (Fig3. (c) and (d)), and the class-specific residuals (Fig3. (e) and (f)) for one given testing sample from class 9 by using DCNR and DSRC on AR. (a) DCNR. (b) DSRC. (c) DCNR. (d) DSRC. (e) DCNR. (f) DSRC.

In addition, the proposed R-DCNR method extends DCNR by using the representation fidelity with l_1-norm instead of l_2-norm. R-DCNR not only has the good properties of DCNR, but also holds the antinoise property that has been proven in [5] and [38]. It should be noted that although the above examples of pair-wise comparisons show the good classification performance, the effectiveness of the proposed methods has been further verified by using extensive experiments in Section V.

B. COMPUTATIONAL COMPLEXITIES OF THE PROPOSED METHODS

In this subsection, we analyze the computational complexities of the proposed DCNR and R-DCNR methods, compared to the related competing methods including CRC [2], SRC [1], [46], CNRC [40], DSRC [37], ProCRC and R-ProCRC [5]. For clearly analyzing their computational complexities, some denotations in Sections II and III are given again as follows: $X \in \mathbb{R}^{m\times N}$, $D \in \mathbb{R}^{N\times N}$, $X_i \in \mathbb{R}^{m\times n}$, $M \in \mathbb{R}^{N\times N}$ and $W_X \in \mathbb{R}^{m\times m}$.

As stated in Section III, since the proposed DCNR method has the closed-form solution $S = (X^TX+\gamma D^TD+2\beta(X^TX+LM)+\lambda I)^{-1}X^Ty$ in Eq. (15), the computational complexity of DCNR is to calculate the solution for a given testing sample y. In DCNR, the computational complexity of X^TX is $O(mN^2)$, the one of D^TD is $O(mN)$ because D^TD can be reformulated as $diag(\| y - x_1 \|_2^2, \ldots, \| y - x_N \|_2^2)$, the one of M is $O(mN^2/L)$ because M can be reformulated as $diag(X_1^TX_1, \ldots, X_L^TX_L)$, the one of $F = (X^TX + \gamma D^TD + 2\beta(X^TX + LM) + \lambda I)^{-1}$ is $O(N^3)$, the one of FX^T is $O(mN^2)$, and finally the one of FX^Ty is $O(mN)$. Thus, the total computational complexity of DCNR for a testing sample is $O(mN^2 + mN + N^3)$.

The solution of R-DCNR is $S = (X^TW_XX + \gamma D^TD + 2\beta(X^TX + LM) + \lambda I)^{-1}X^TW_Xy$ that should be iteratively updated until the maximum iteration T is reached. Compared to DCNR, here we can only calculate the computational complexity of X^TW_XX to obtain the total computational complexity of R-DCNR. The computation complexity of X^TW_XX is $O(Nm^2+mN^2)$, and then the one of S is $O(Nm^2+mN^2+mN+N^3)$ in terms of the computational complexity of DCNR. Since W_X and S are updated T times, thus the computational complexity of R-DCNR is $O(TNm^2 + TmN^2 + TmN + TN^3)$.

We also give the computational complexities of CRC, CNRC, DSRC, ProCRC, R-ProCRC and SRC, in comparisons with the proposed methods. The computational complexity of CRC is to calculate the solution $S = (X^TX + \lambda I)^{-1}X^Ty$ with $O(mN^2 + mN + N^3)$ [2]. The computational complexity of SRC is to calculate the solution by minimizing the l_1-norm problem using the dual augmented lagrangian method with $O(Tm^2 + TmN)$ [1], [46]. The computational complexity of DSRC is to calculate the solution $S = ((1 + 2\gamma)X^TX + 2\lambda LM)^{-1}X^Ty$ with $O(mN^2 + mN + N^3)$ [37]. The computational complexity of CNRC is to calculate the solution $S = (X^TX+\gamma D^TD+\lambda I)^{-1}X^Ty$ with $O(mN^2+mN+N^3)$ [40]. The computational complexity of ProCRC is to calculate the solution $S = (X^TX+\frac{\gamma}{L}\sum_{c=1}^{L}(\bar{X}_c')^T\bar{X}_c'+\lambda I)^{-1}X^Ty$ with $O(mN^2+mN+N^3)$, where $X_c' = [0, \cdots, X_c, \cdots, 0] \in \mathbb{R}^{m\times N}$ and $\bar{X}_c' = X - X_c'$ [5]. The computational complexity of R-ProCRC is to calculate the solution $S = (X^TW_XX + \frac{\gamma}{L}\sum_{c=1}^{L}(\bar{X}_c')^T\bar{X}_c'+\lambda I)^{-1}X^Ty$ with $O(TNm^2+TmN^2+TmN+TN^3)$ [5]. The computational complexities of the comparative methods are summarized in Table 1. Therefore, we can clearly observe that the proposed methods have the closed-solutions with the similar computational complexities as the related variants of CRC.

TABLE 1. The computational complexities of the competing methods.

Methods	Computational complexity
DCNR	$O(mN^2 + mN + N^3)$
R-DCNR	$O(TNm^2 + TmN^2 + TmN + TN^3)$
CRC	$O(mN^2 + mN + N^3)$
CNRC	$O(mN^2 + mN + N^3)$
DSRC	$O(mN^2 + mN + N^3)$
ProCRC	$O(mN^2 + mN + N^3)$
R-ProCRC	$O(TNm^2 + TmN^2 + TmN + TN^3)$
SRC	$O(Tm^2 + TmN)$

V. EXPERIMENTS

In this section, to verify the effectiveness of the proposed DCNR and R-DCNR methods, extensive experiments on seven public face data sets including AR, FEI, GT, IMM, LFW, ORL and PIE are conducted by comparing both DCNR and R-DCNR to CRC [2], SRC [1], CNRC [40], DSRC [37], ProCRC and R-ProCRC [5] in terms of the recognition accuracy.

A. DATA SETS

In this subsection, we briefly describe the used seven face data sets in the experiments. The AR face data set[1] has 4000 image samples taken from 126 persons. The subset of AR is used and has 100 persons including 50 females and 50 males. Each person has 14 image samples taken by different expressions and illuminations. Fig. 4(a) shows the example of the image samples from two persons in AR. The Georgia Tech (GT) face data set[2] has 750 frontal or titled face images from 50 persons. Each person has 15 image samples taken by varying the positions, expressions and lighting conditions. Fig. 4(b) shows the example of the image samples from two persons in GT. The IMM face data set[3] has 240 image samples from 40 persons including 33 males and 7 females. Each person has 6 image samples taken by the rotated variations from 30 left degrees to 30 right degrees or lighting variations. Fig. 4(c) shows the example of the image samples from two persons in IMM. The ORL face data set[4] has 400 image samples from 40 persons. Each person has 10 image samples taken by the various illuminations and facial expressions. Fig. 4(d) shows the example of the image samples from two persons in ORL. The FEI face data set[5] has 2800 image samples from 200 persons including 100 males and 100 females. Each person has 14 samples taken in an upright frontal position

[1] http://www2.ece.ohio-state.edu/∼aleix/ ARdatabase.html
[2] http://www. anefian.com/research/face_reco.htm
[3] http://www.imm.dtu.dk/∼aam/ datasets/datasets.html
[4] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[5] http://fei.edu.br/ cet/facedatabase.html

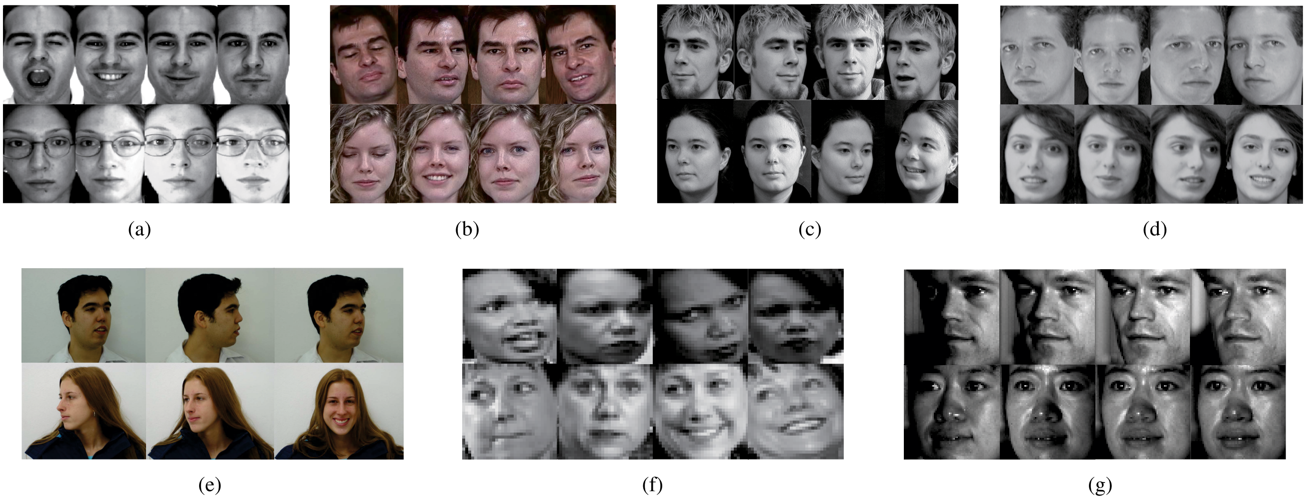


FIGURE 4. **The examples of some image face samples for two subjects on each face data set. (a) AR. (b) GT. (c) IMM. (d) ORL. (e) FEI. (f) LFW. (g) PIE.**

with profile rotation of up to about 180 degrees. Fig. 4(e) shows the example of the image samples from two persons in FEI. The LFW (Labeled Faces in the Wild) face data set[6] has more than 13000 face images taken from the web for unconstrained face recognition. We use a subset of LFW with 1251 image samples from 86 persons. Each person has about 11-20 samples [11]. Fig. 4(f) shows the example of the image samples from two persons in LFW. The PIE face data set[7] has 41386 image samples from 68 persons. We use a subset of PIE with 1632 samples from 68 persons, each of which has 24 images under various view points, illuminations and expressions. Fig. 4(g) shows the example of the image samples from two persons in PIE.

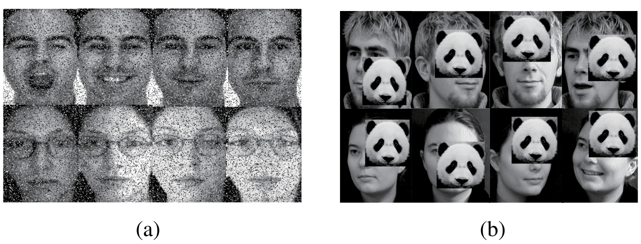


FIGURE 5. **The examples of the testing image samples with random corruptions on AR and block occlusions on IMM. (a) AR. (b) IMM.**

In addition, we add random corruptions and block occlusions into all the testing samples on the AR, GT, ORL and IMM data sets, in order to verify the robustness of the proposed R-DCNR. Fig. 5(a) shows the example of the testing image samples with the 30% random corruption rate on AR and Fig. 5(b) shows the example of the testing image samples with the 30% block occlusions on IMM. Note that the sizes of all the image samples are resized to 32×32 pixels in the experiments.

[6]http://vis-www.cs.umass.edu/lfw/
[7]http://www.flintbox.com/public/project/4742/

B. EXPERIMENT 1

In this subsection, the experiment is to evaluate the classification performance of the proposed DCNR method on seven face data sets without corruptions and occlusions, compared to CRC, SRC, CNRC, DSRC and ProCRC. In the experiments, each face data set is randomly divided into the training set and testing set several times. And then the average classification results and the corresponding standard deviations on each data set can be obtained by each competing method. The numbers t of the training samples on these face data sets are preseted in the ranges of 3 to 9 on AR with step 1, 3 to 9 on GT with step 1, 2 to 5 on IMM with step 1, 1 to 6 on ORL with step 1, 1 to 9 on FEI with step 1, 1 to 6 on LFW with step 1, and 4 to 14 on PIE with step 2. The average classification accuracies of the competing methods with varying the different numbers of training samples are displayed in Fig. 6. We can obviously observe that the classification accuracies of each RBC method quickly ascend with an increase of the numbers of training samples. Furthermore, we can see that the proposed DCNR almost performs best among the competing methods at each number of the training samples on each face data set. Through the comparisons of the classification performance of the CRC methods in Fig. 6, it means that the locality term and discrimination terms integrated in the proposed DCNR method are very meaningful for improving the power of the pattern discrimination.

To particularly display the good classification performance of the proposed DCNR, we provide the average recognition accuracies of the competing methods in Table 2. The average classification results in Table 2 correspond to them in Fig. 6 when the numbers t of the training samples on these face data sets are set as follows: $t = 4, 6$ on AR, $t = 5, 8$ on GT, $t = 4, 5$ on IMM, $t = 3, 4$ on ORL, $t = 2, 5$ on FEI, $t = 5, 6$ on LFW, and $t = 4, 6$ on PIE. Note that the best classification results among these competing methods on each face data set are indicated in bold. It is obvious

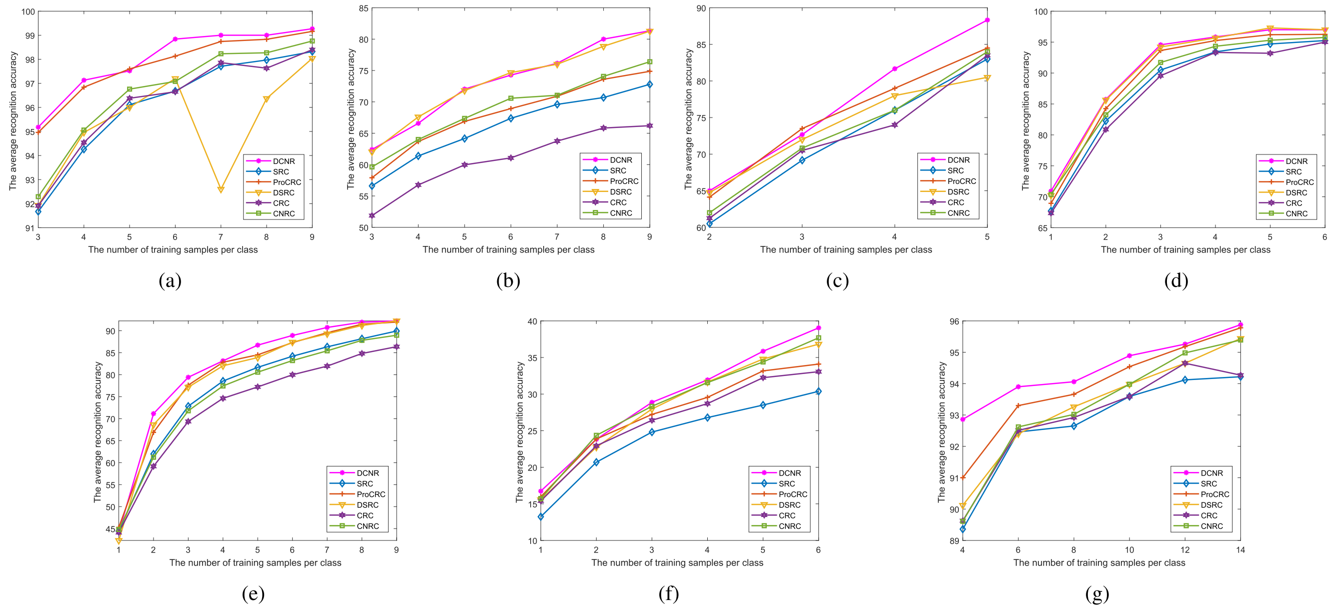


FIGURE 6. The average recognition accuracies (%) of each method with varying differen numbers of the training samples on all the face data sets. (a) AR. (b) GT. (c) IMM. (d) ORL. (e) FEI. (f) LFW. (g) PIE.

TABLE 2. The average classification accuracies (%) of each method with the corresponding standard deviations on all the face data sets.

Data set	t	CRC	SRC	ProCRC	CNRC	DSRC	DCNR
AR	4	94.54±0.50	94.26±0.57	96.84±0.67	95.06±0.76	94.96±0.93	**97.13±0.49**
	6	96.65±0.56	96.68±0.74	98.13±0.42	97.08±0.57	97.20±0.83	**98.84±0.57**
GT	5	59.96±1.13	64.16±1.31	66.88±1.35	67.36±0.55	71.80±1.66	**72.04±1.34**
	8	65.83±1.32	70.69±1.08	73.60±1.45	74.06±0.89	78.86±1.29	**80.00±1.38**
IMM	4	74.00±3.35	76.00±3.11	79.00±3.58	76.00±2.98	78.00±4.01	**81.67±1.44**
	5	83.50±3.35	83.00±3.26	84.50±2.74	84.00±3.79	80.50±6.47	**88.33±2.89**
ORL	3	89.57±2.69	90.50±2.42	93.64±1.27	91.71±2.77	94.21±1.64	**94.57±1.81**
	4	93.33±1.02	93.42±1.00	95.25±0.76	94.33±0.37	95.67±1.20	**95.83±0.66**
FEI	2	59.16±1.61	61.97±1.19	66.88±1.41	61.23±1.13	68.63±1.31	**71.14±0.63**
	5	77.22±1.81	81.67±1.39	84.52±1.25	80.57±2.02	83.89±1.55	**86.72±0.97**
LFW	5	32.24±0.70	28.50±1.48	33.17±1.69	34.39±1.83	34.79±0.46	**35.85±1.13**
	6	33.06±1.19	30.38±0.57	34.10±1.29	37.69±2.36	36.82±2.70	**39.05±0.89**
PIE	4	89.63±0.73	89.36±0.94	91.00±0.61	89.61±0.74	90.12±0.63	**92.86±0.61**
	6	92.51±0.10	92.46±0.10	93.30±0.14	92.62±0.09	92.40±0.14	**93.90±0.23**

from Table 2 that the proposed DCNR almost significantly outperforms CRC, SRC, CNRC, DSRC and ProCRC under these given numbers of the training samples on each face data set. Consequently, it can be concluded that the superior classification performance of the proposed DCNR over the other competing RBC methods benefits from the locality and discrimination constraints on the representation.

C. EXPERIMENT 2

In this subsection, we investigate the robust classification performance of the proposed R-DCNR method in comparisons with the proposed DCNR, CRC, SRC, CNRC, DSRC, ProCRC and R-ProCRC under the condition of all the testing samples with random corruptions on the AR, GT, ORL and IMM face data sets. It should be noted that the robust ProCRC (R-ProCRC) for image classification is employed to comparatively verify the proposed R-DCNR, because both R-ProCRC and R-DCNR adopt the l_1-norm based representation fidelities. In the experiments, the first t samples of each class on each face data set are chosen as the training samples, and the remaining ones are the testing samples. The numbers t are set as follows: $t = 4, 7$ on AR, $t = 4, 7$ on GT, $t = 3, 5$ on IMM, $t = 4, 7$ on ORL. For the random corruptions, 10%, 20%, 30%, 40% and 50% of the pixels of each testing sample are corrupted by using the uncertain grayscale values between 0 and 255.

The recognition accuracies of each method with varying the random corruption rates on AR, GT, ORL and IMM are

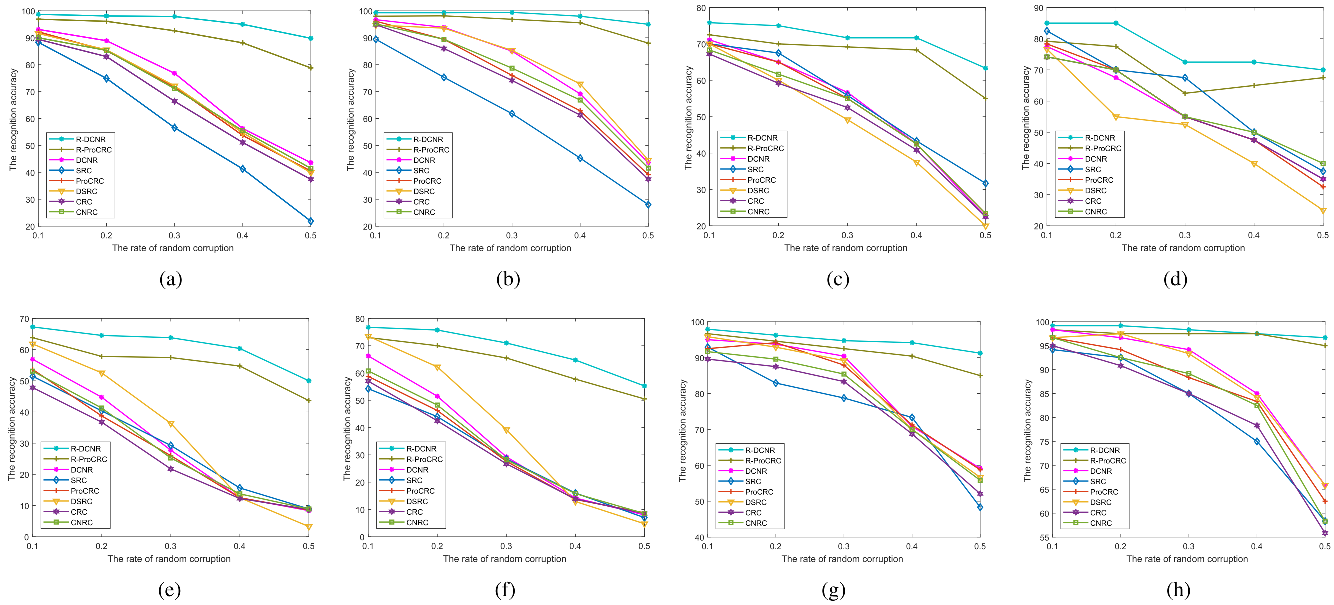


FIGURE 7. The recognition accuracies (%) of the competing methods on the AR, IMM, GT, ORL face data sets with varying the random corruption rates. (a) AR ($t = 4$). (b) AR ($t = 7$). (c) IMM ($t = 3$). (d) IMM ($t = 5$). (e) GT ($t = 4$). (f) GT ($t = 7$). (g) ORL ($t = 4$). (h) ORL ($t = 7$).

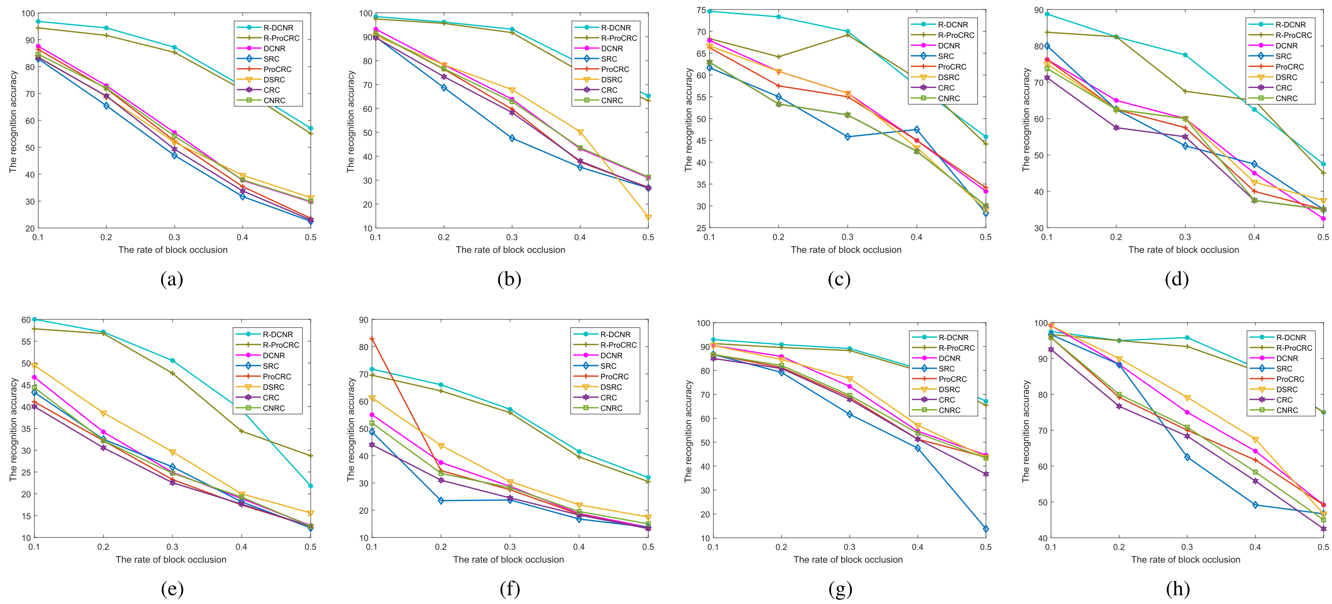


FIGURE 8. The recognition accuracies (%) of the competing methods on the AR, IMM, GT, ORL face data sets with varying the block occlusion rates. (a) AR ($t = 4$). (b) AR ($t = 7$). (c) IMM ($t = 3$). (d) IMM ($t = 5$). (e) GT ($t = 4$). (f) GT ($t = 7$). (g) ORL ($t = 4$). (h) ORL ($t = 7$).

shown in Fig. 7. It can be seen from Fig. 7 that the proposed R-DCNR performs best among the competing methods under the condition of the random corruptions. We can observe that the classification accuracies of both R-ProCRC and R-DCNR with l_1-norm based representation fidelities are better than DCNR, CRC, SRC, CNRC, DSRC and ProCRC with l_2-norm based representation fidelities. It can also be seen that the recognition accuracies of each method decrease with an increase of the ratios of the random corruptions, but the proposed R-DCNR is very less sensitive to the random corruptions than the other methods. In addition, the proposed DCNR often robustly perform better than the other methods with with l_2-norm based representation fidelities. Thus, the experimental results in Fig. 7 show that proposed R-DCNR obtains the robust classification performance by simultaneously using the l_1-norm based representation fidelity, the locality term and discrimination term.

D. EXPERIMENT 3

In this subsection, the robustness of the proposed R-DCNR method is further verified on the AR, GT, ORL and IMM face data sets in which the testing samples are added by the block occlusions. Under the condition of the block occlusions, we also compare R-DCNR to DCNR, CRC, SRC, CNRC, DSRC, ProCRC and R-ProCRC. For the block occlusions, 10%, 20%, 30%, 40% and 50% of the size of each testing sample are replaced by the head image of a panda in arbitrary positions of the testing image sample. Just as in Subsection V-C, the first t samples of each class are selected as the training samples and the rest of samples of each class are the testing samples. And the values of t on AR, GT, ORL and IMM are the same as in Subsection V-C in the experiments.

The recognition accuracies of the competing methods with the different rates of the block occlusions on AR, GT, ORL and IMM are displayed in Fig. 8. We can see that the proposed R-DCNR almost achieves the best performance among the competing methods with the different block occlusions. Moreover, the classification performance of each method degrades with an increase of the block occlusion rates. Meantime, the superior classification performance of both R-DCNR and R-ProCRC over CRC, SRC, CNRC, DSRC, ProCRC and DCNR means that the method with the l_1-norm based representation fidelity in Fig. 8 is more robust to the noises than the one with the l_2-norm based representation fidelity. Hence, the proposed R-DCNR is effective and robust for classification.

VI. CONCLUSION

In the RBC methods, the classification performance can be improved by fully considering the locality and discrimination information of data. In this article, based on both CNRC and DSRC, we proposed a novel discriminative collaborative neighbor representation method (DCNR) by appropriately integrating the locality term from CNRC and the discrimination term from DSRC. To further improve the classification performance of the proposed method under the condition of data with noises, we use the power of discrimination from the l_1-norm based representation fidelity to design the robust DCNR method (R-DCNR). Namely, the proposed DCNR is extended to R-DCNR by using the l_1-norm based representation fidelity instead of the l_2-norm based representation fidelity. To demonstrated the proposed DCNR, we conduct the experiments on seven public face data sets, and compare DCNR to CRC, SRC, CNRC, DSRC and ProCRC. Meanwhile, to verify the the proposed R-DCNR, the experiments have been done on four face data sets with the random corruptions and the block occlusions in comparisons with DCNR, CRC, SRC, CNRC, DSRC, ProCRC and R-ProCRC. The experimental results show the effectiveness of the proposed DCNR and the robustness of R-DCNR in pattern classification. In the future works, we will plan to apply the proposed methods to the practical object recognition with some deep learning methods.

REFERENCES

[1] J. Wright, A. Y. Yang, A. Ganesh, 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.

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

[3] 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.

[4] W. Deng, J. Hu, and J. Guo, "Face recognition via collaborative representation: Its discriminant nature and superposed representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2513–2521, Oct. 2018.

[5] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2950–2959.

[6] J. Gou, Y. Xu, D. Zhang, Q. Mao, L. Du, and Y. Zhan, "Two-phase linear reconstruction measure-based classification for face recognition," *Inf. Sci.*, vols. 433–434, pp. 17–36, Apr. 2018.

[7] W. Ou *et al.*, "Robust discriminative nonnegative dictionary learning for occluded face recognition," *Pattern Recognit. Lett.*, vol. 107, pp. 41–49, May 2018.

[8] W. Li, Q. Du, F. Zhang, and W. Hu, "Hyperspectral image classification by fusing collaborative and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4178–4187, Sep. 2016.

[9] W. Li, J. Liu, and Q. Du, "Sparse and low-rank graph for discriminant analysis of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4094–4105, Jul. 2016.

[10] R. Lan and Y. Zhou, "An extended probabilistic collaborative representation based classifier for image classification," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 1392–1397.

[11] 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.*, vol. 28, no. 2, pp. 278–293, Feb. 2017.

[12] 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, 2015.

[13] Y. Xu, Z. Li, J. Yang, and D. Zhang, "A survey of dictionary learning algorithms for face recognition," *IEEE Access*, vol. 5, pp. 8502–8514, 2017.

[14] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[15] G. Zhong, L. Wang, X. Ling, and J. Dong, "An overview on data representation learning: From traditional feature learning to recent deep learning," *J. Finance Data Sci.*, vol. 2, no. 4, pp. 265–278, 2016.

[16] H. Cheng, Z. Liu, L. Yang, and X. Chen, "Sparse representation and learning in visual recognition: Theory and applications," *Signal Process.*, vol. 93, no. 6, pp. 1408–1425, 2013.

[17] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognit.*, vol. 43, no. 1, pp. 331–341, 2010.

[18] C.-Y. Lu, H. Min, J. Gui, L. Zhu, and Y.-K. Lei, "Face recognition via weighted sparse representation," *J. Vis. Commun. Image Represent.*, vol. 24, no. 2, pp. 111–116, 2013.

[19] Z. Fan, M. Ni, Q. Zhu, and E. Liu, "Weighted sparse representation for face recognition," *Neurocomputing*, vol. 151, no. 1, pp. 304–309, 2015.

[20] X. Tang, G. Feng, and J. Cai, "Weighted group sparse representation for undersampled face recognition," *Neurocomputing*, vol. 145, pp. 402–415, Dec. 2014.

[21] C.-P. Wei, Y.-W. Chao, Y.-R. Yeh, and Y.-C. F. Wang, "Locality-sensitive dictionary learning for sparse representation based classification," *Pattern Recognit.*, vol. 46, no. 5, pp. 1277–1287, 2013.

[22] Y. Zhan, J. Liu, J. Gou, and M. Wang, "A video semantic detection method based on locality-sensitive discriminant sparse representation and weighted KNN," *J. Vis. Commun. Image Represent.*, vol. 41, pp. 65–73, Nov. 2016.

[23] J. Liu, J. Gou, Y. Zhan, and Q. Mao, "Discriminative self-adapted locality-sensitive sparse representation for video semantic analysis," *Multimedia Tools Appl.*, vol. 77, no. 21, pp. 29143–29162, 2018, doi: 10.1007/s11042-018-6090-6.

[24] J. Wang, C. Lu, M. Wang, P. Li, S. Yan, and X. Hu, "Robust face recognition via adaptive sparse representation," *IEEE Trans. Cybern.*, vol. 44, no. 22, pp. 2368–2378, Dec. 2014.

[25] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, 2014.

[26] J. Li and C.-Y. Lu, "A new decision rule for sparse representation based classification for face recognition," *Neurocomputing*, vol. 116, pp. 265–271, Sep. 2013.

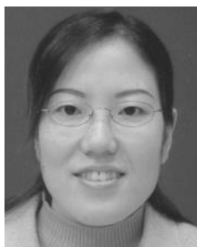
[27] H. Ma, J. Gou, X. Wang, J. Ke, and S. Zeng, "Sparse coefficient-based k-nearest neighbor classification," *IEEE Access*, vol. 5, pp. 16618–16634, 2017.
[28] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang. (2012). "Collaborative representation based classification for face recognition." [Online]. Available: https://arxiv.org/abs/1204.2358
[29] Y. Wu, W. Li, M. Mukunoki, M. Minoh, and S. Lao, "Discriminative collaborative representation for classification," in *Computer Vision—ACCV* (Lecture Notes in Computer Science), vol. 9006. Cham, Switzerland: Springer, 2015.
[30] Y. Lei, Y. Guo, M. Hayat, M. Bennamoun, and X. Zhou, "A two-phase weighted collaborative representation for 3D partial face recognition with single sample," *Pattern Recognit.*, vol. 52, pp. 218–237, Apr. 2016.
[31] Z. Liu, J. Pu, M. Xu, and Y. Qiu, "Face recognition via weighted two phase test sample sparse representation," *Neural Process. Lett.*, vol. 41, no. 1, pp. 43–53, 2015.
[32] J. Gou, Y. Zhan, X. Shen, Q. Mao, and L. Wang, "Two-phase representation based classification," in *Advances in Multimedia Information Processing—PCM 2015* (Lecture Notes in Computer Science), vol. 9314, Y. S. Ho, J. Sang, Y. Ro, J. Kim, and F. Wu, Eds. Cham, Switzerland: Springer, 2015.
[33] H. Yuan, X. Li, F. Xu, Y. Wang, L. L. Lai, and Y. Y. Tang, "A collaborative-competitive representation based classifier model," *Neurocomputing*, vol. 275, pp. 627–635, Jan. 2018.
[34] H. Chi, H. Xia, L. Zhang, C. Zhang, and X. Tang, "Competitive and collaborative representation for classification," *Pattern Recognit. Lett.*, to be published, doi: 10.1016/j.patrec.2018.06.019.
[35] D. M. Vo and S.-W. Lee, "Robust face recognition via hierarchical collaborative representation," *Inf. Sci.*, vol. 432, pp. 332–346, Mar. 2018.
[36] Y. Xu, X. Li, J. Yang, Z. Lai, and D. Zhang, "Integrating conventional and inverse representation for face recognition," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1738–1746, Oct. 2014.
[37] Y. Xu, Z. Zhong, J. Yang, J. You, and D. Zhang, "A new discriminative sparse representation method for robust face recognition via l_2 regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2233–2242, Oct. 2017.
[38] S. Zeng, J. Gou, and L. Deng, "An antinoise sparse representation method for robust face recognition via joint l_1 and l_2 regularization," *Expert Syst. Appl.*, vol. 82, pp. 1–9, Oct. 2017.
[39] R. Timofte and L. Van Gool, "Adaptive and weighted collaborative representations for image classification," *Pattern Recognit. Lett.*, vol. 43, pp. 127–135, Jul. 2014.
[40] J. Waqas, Z. Yi, and L. Zhang, "Collaborative neighbor representation based classification using l_2-minimization approach," *Pattern Recognit. Lett.*, vol. 34, no. 2, pp. 201–208, 2013.
[41] Z. Li, Q. Zhu, B. Xie, J. Cao, and J. Zhang, "A collaborative neighbor representation based face recognition algorithm," *Math. Problems Eng.*, vol. 2013, Aug. 2013, Art. no. 373858.
[42] H. Chi, H. Xia, X. Tang, Y. Zhang, and X. Xia, "Supervised neighborhood regularized collaborative representation for face recognition," *Multimedia Tools Appl.*, vol. 77, no. 22, pp. 29509–29529, 2018.
[43] T. Jin, Z. Liu, Z. Yu, X. Min, and L. Li, "Locality preserving collaborative representation for face recognition," *Neural Process. Lett.*, vol. 45, no. 3, pp. 967–979, 2017.
[44] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
[45] R. Chartrand and W. Yin, "Iterative reweighted algorithms for compressive sensing," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Mar./Apr. 2008, pp. 3869–3872.
[46] A. Y. Yang, S. S. Sastry, A. Ganesh, 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.

JIANPING GOU received the Ph.D. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2012. He is currently an Associate Professor with the School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang, China. His current research interests include pattern recognition and machine learning.

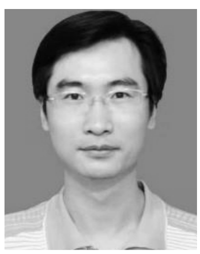
LEI WANG is currently pursuing the master's degree with the School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang, China. His research interests include image classification and machine learning.

ZHANG YI (F'–) received the Ph.D. degree in mathematics from the Institute of Mathematics, Chinese Academy of Science, Beijing, China, in 1994. He is currently a Professor and the Dean with the College of Computer Science, Sichuan University, Chengdu, China. His current research interests include neural networks and big data.

JIANCHENG LV received the Ph.D. degree in computer science and engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2006. He is currently a Professor with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu. His research interests include neural networks, machine learning, and big data.



QIRONG MAO received the Ph.D. degree in computer application technology from Jiangsu University, Zhenjiang, China, in 2009. She is currently a Professor with the School of Computer Science and Communication Engineering, Jiangsu University. Her research interests include affective computing, pattern recognition, and multimedia analysis.



YUN-HAO YUAN received the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology (NUST), China, in 2013. He is currently an Associate Professor with the Department of Computer Science and Technology, College of Information Engineering, Yangzhou University. His research interests include pattern recognition, machine learning, multimedia search, and information fusion. He received two national scholarships from the Ministry of Education, China, and the Outstanding Ph.D. Thesis Award from NUST.