Pattern Recognition 66 (2017) 129–143

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Learning robust and discriminative low-rank representations for face recognition with occlusion

Guangwei Gao[a,b,c,*], Jian Yang[d], Xiao-Yuan Jing[b,e], Fumin Shen[f], Wankou Yang[g], Dong Yue[a,b]

[a] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China
[b] Jiangsu Engineering Laboratory of Big Data Analysis and Control for active distribution network, Nanjing University of Posts and Telecommunications, Nanjing, China
[c] Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University), Fuzhou, China
[d] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
[e] School of Automation, Nanjing University of Posts and Telecommunications, Nanjing, China
[f] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China
[g] School of Automation, Southeast University, Nanjing, China

ARTICLE INFO

Keywords:
Face recognition
Low-rank matrix recovery
Nuclear norm

ABSTRACT

For robust face recognition tasks, we particularly focus on the ubiquitous scenarios where both training and testing images are corrupted due to occlusions. Previous low-rank based methods stacked each error image into a vector and then used L_1 or L_2 norm to measure the error matrix. However, in the stacking step, the structure information of the error image can be lost. Depart from the previous methods, in this paper, we propose a novel method by exploiting the low-rankness of both the data representation and each occlusion-induced error image simultaneously, by which the global structure of data together with the error images can be well captured. In order to learn more discriminative low-rank representations, we formulate our objective such that the learned representations are optimal for classification with the available supervised information and close to an ideal-code regularization term. With strong structure information preserving and discrimination capabilities, the learned robust and discriminative low-rank representation (RDLRR) works very well on face recognition problems, especially with face images corrupted by continuous occlusions. Together with a simple linear classifier, the proposed approach is shown to outperform several other state-of-the-art face recognition methods on databases with a variety of face variations.

1. Introduction

Object classification is a fundamental problem in pattern recognition community [1–16]. Due to the advantages of non-intrusive natural and high uniqueness, face recognition has been an interesting and active topic in many ubiquitous biometrics applications [17–32], such as surveillance, human machine interaction, access control, photo album management in social media. In the past few decades, regression based face recognition approaches have led to state-of-the-art performance [33–44]. The representative ones are sparse representation-based classification (SRC) [33] and linear regression-based classification (LRC) [34]. SRC solves an L_1-norm minimization problem for a test input by deriving the sparse coefficients for the training data, and then uses the coding coefficients associated with each class to calculate the distance from the query sample to each class. Many following works of SRC have been reported for vision problems, e.g., super-resolution [45], facial expression recognition [46] and visual tracking [47], etc. On the other side, Naseem et al. [34] proposed LRC for face recognition. Based on an assumption that samples from a specific object class lie on a linear subspace, LRC represents a test image as a linear combination of training images of each class. Yang et al. [48] gave an insight into SRC and sought reasonable supports for its effectiveness. They thought that the L_1-regularizer has two properties, sparseness and closeness. Sparseness determines a small number of nonzero representation coefficients and closeness makes the nonzero representation coefficients concentrating on the training samples having the same class label as the test sample. Based on the discussion about the working mechanism of SRC, Zhang et al. [36] demonstrated that the role of collaboration between classes in representing a query image is more important than that of the sparsity constraints. In their work, a collaborative representation-based classification (CRC) model is presented with a squared L_2-regularization which achieves competitive

* Corresponding author at: Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, No. 9 Wenyuan Road, Nanjing 210023, China.
E-mail address: csggao@gmail.com (G. Gao).

http://dx.doi.org/10.1016/j.patcog.2016.12.021
Received 8 July 2016; Received in revised form 19 December 2016; Accepted 20 December 2016
Available online 21 December 2016
0031-3203/

performance in terms of accuracy but with significantly lower complexity than that of the sparse representation method.

It is worth noting that most previous works assume that both training and testing images are taken under a well-controlled setting (i.e., under reasonable illumination, pose, variations without occlusion or disguise). Their performance will be degraded when the images are contaminated. To further assess the robustness of the proposed algorithm, test images are considered to be corrupted due to occlusion or disguise in literatures. By introducing an identity matrix I as a dictionary to code the outliers (e.g., corrupted or occluded pixels), SRC [33] exhibits excellent robustness and produces promising performance. However, SRC is not robust to contiguous occlusion such as sunglasses and scarf, since the occlusion level has gone beyond the breakdown point determined by the algorithm. By modeling the sparse coding as a sparsity-constrained robust regression problem, Yang et al. [49] modified the SRC framework for handling outliers such as occlusions in face recognition. He et al. [50] unified the algorithms for error correction and detection by using the additive and the multiplicative forms respectively, and established a half-quadratic framework to solve the problem of robust sparse representation. Yang et al. [51–54] used nuclear norm to describe the structural characteristics of error image and proposed nuclear norm based matrix regression and matrix decomposition model, which is robust for face recognition with occlusion and illumination changes and for occluded image recovery. Unfortunately, the performance of the above methods will be degraded if both training and testing images are corrupted, since none of them take the possible corruption in training images into account.

Low-rank matrix recovery, which determines a low-rank matrix from corrupted input data, has been successfully applied to applications including salient object detection [55], background subtraction [56], visual tracking [57] and image denoising [58]. Recently, this technique has been applied for multi-class classification. Chen et al. [59,60] used robust principle component analysis (RPCA) [61,62] to first remove noise from the training data class by class. Then traditional PCA plus SRC were employed for feature extraction and recognition. Motivated by low-rank matrix recovery, Ma et al. [63] presented a discriminative low-rank dictionary learning algorithm to learn a low-rank dictionary class by class for sparse representation-based face recognition. Although the above matrix recovery based methods have received promising results in case that both the training and testing images are corrupted, they remove noise from training samples class by class. This process is computationally expensive when the number of classes is large. On the other hand, the class-by-class strategy ignores the correlation between different classes.

As an extension of RPCA, low-rank representation (LRR) [64] was originally presented to segment subspace from a union of multiple subspaces. LRR can capture the global structure of data and perform robust subspace segmentation from the corrupted data. Later, Liu et al. [65] proposed a latent low-rank representation (LatLRR) algorithm to integrate subspace segmentation and feature extraction into a unified framework. LatLRR is able to robustly extract salient features from corrupted data for classification. Similarly, Yin et al. [66] proposed a double low-rank matrix recovery method by considering the recovery of row space and column space information simultaneously. In literatures [65,66], one of the low-rank matrices was used as projection function to extract salient features for classification. By introducing an idea regularization term into the objective of LRR, Zhang et al. [67] proposed to learn a structured low-rank representation from contaminated training samples for classification. A discriminative low-rank representation with respect to the constructed dictionary is obtained and a simple yet powerful linear multi-classifier is performed on this representation for classification tasks.

In previous low-rank based methods, each error image is stacked into a vector and all the representation residuals form an error matrix, which is characterized by the L_1 or L_2 norm. However, after this stacking step, the structure information of the image may be ignored. Meanwhile, characterizing the representation error pixel by pixel individually may neglect the whole structure of the error image. To capture the low-rank structure information in the error image, particularly with occlusion-induced errors, we introduce the low-rank assumption to characterize the representation and error term simultaneously: the representations form a low-rank matrix, which can capture the global structure of data and encode the pairwise affinities between data vectors; each error image is also a low-rank matrix, which can directly characterize the holistic structure of the occlusion-induced error image individually. Label information from the training data is then incorporated by adding an ideal-code regularization term to the objective function of our model. In addition, the classification error constraint is included to make the learned representation optimal for final classification purpose. With the above low-rank assumption, ideal-code regularization term and classification error constraint, we are able to learn a robust and discriminative low-rank representation (RDLRR) for face recognition with occlusions. RDLRR aims to reveal the global structure of data in vector representation space and remove the low-rank error in original image space simultaneously. After obtaining the robust and discriminative representation, we also use a simple yet powerful linear multi-classifier for classification tasks. Our extensive experiments will verify the effectiveness and robustness of the proposed method.

The reminder of the paper is organized as follows. Section 2 briefly reviews related works on low-rank matrix recovery for classification. In Section 3, we present our proposed algorithm based on low-rank representation, including the optimization details. Additionally, we also provide the complexity analysis and convergence analysis in this section. Section 4 evaluates the performance of the proposed methods on several commonly used face recognition databases. Section 5 concludes this paper.

2. Low-rank matrix recovery for classification

In this section, we first review two popular low-rank matrix recovery models, including robust principal component analysis (RPCA) and low-rank representation (LRR). Then, we introduce the classification scheme based on robust representation.

2.1. Low-rank matrix recovery

Principal component analysis (PCA) [68] is a well-known dimension reduction technique for image reconstruction and classification purpose. Despite its efficiency and effectiveness, PCA is known to be sensitive to errors with large magnitudes. When the data is contaminated (occlusion or disguise), the reconstruction and classification performance of PCA deteriorates rapidly. A number of approaches have been proposed in literatures to address this problem, including the introduction of L_1-norm variance [69], and low-rank matrix recovery [61,62].

Low-rank matrix recovery aims at decomposing data matrix $\mathbf{X}$ into two matrixes $\mathbf{A}$ and $\mathbf{E}$, where $\mathbf{A}$ is a low-rank matrix and $\mathbf{E}$ is the corresponding sparse error matrix. To recovery the low-rank approximation $\mathbf{A}$ from $\mathbf{X}$, Low-rank matrix recovery needs to solve the following minimization problem:

$$\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_l \quad s.t. \quad \mathbf{X} = \mathbf{A} + \mathbf{E}, \tag{1}$$

where λ is a parameter that controls the sparsity of the noise matrix $\mathbf{E}$, $||\mathbf{A}||_*$ is the nuclear norm (i.e., the sum of the singular values) of $\mathbf{A}$, $||\mathbf{E}||_1$ is the L_1-norm (i.e., the sum of the absolute values of each entry) of $\mathbf{E}$. As proved in [61], under some suitable assumptions, it is possible to recover both the low-rank and the sparse components exactly by solving a convenient convex program. Model (1) assumes that all data vectors in $\mathbf{X}$ are coming from a single subspace. Chen et al. [59, 60]

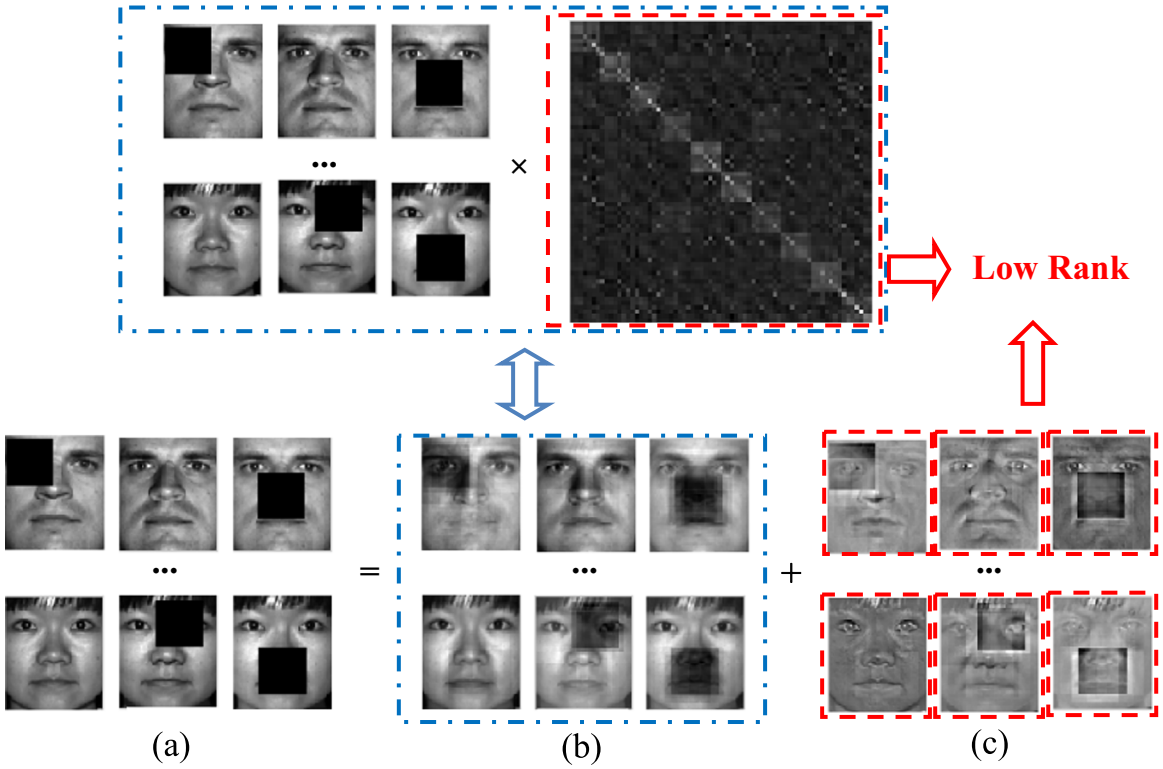

Fig. 1. Illustration of image decomposition with low-rank assumption. (a) Original images, (b) Recovered images, (c) Error images.

used this model to remove noise from training samples class by class, this process can be computationally expensive for large number of classes.

In many cases such as face recognition, the underlying dataset is a union of multiple subjects. Samples of one subject may be drawn from the same subspace, while samples of different subjects are from different subspaces. LRR [64] seeks the lowest-rank representation among all the candidates that represent all samples as the linear combination of the bases in a dictionary. A more general rank minimization problem is formulated as:

$$\min_{Z,D,E} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_l \qquad s.\,t. \quad \mathbf{X} = \mathbf{DZ} + E, \tag{2}$$

where $\mathbf{D}$ is a dictionary that linearly spans the data space and $\mathbf{Z}$ is the lowest-rank representation of data $\mathbf{X}$ over dictionary $\mathbf{D}$. By choosing the dictionary $\mathbf{D} = \mathbf{I}$, Eq. (2) degenerates to Eq. (1). So LRR can be regarded as a generalization of RPCA. LRR can capture the global structure of data, giving a more effective tool for robust subspace segmentation and other applications.

2.2. Classification

An easy yet powerful linear classifier is used for classification tasks [67,70]. Denote by $\mathbf{Z}$ the low-rank representation of training data $\mathbf{X}$ and by $\mathbf{Z_{test}}$ the low-rank representation of test data $\mathbf{X_{test}}$ over the dictionary $\mathbf{D}$ respectively. The multivariate ridge regression is used to train a linear classifier $\hat{\mathbf{W}}$:

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \|\mathbf{H} - \mathbf{WZ}\|_2^2 + \tau\|\mathbf{W}\|_2^2, \tag{3}$$

where $\mathbf{H}$ is the class label matrix of data $\mathbf{X}$ and $\tau > 0$ is a parameter. Since Eq. (3) is a standard regression model, we can yield its close-form solution by $\hat{\mathbf{W}} = H\mathbf{Z}^T(\mathbf{ZZ}^T + \tau\mathbf{I})^{-1}$. Then label for test sample representation z_i (the ith column in $\mathbf{Z_{test}}$) is given by:

Algorithm 1:. Low-rank matrix recovery for classification.

Input: Training data $\mathbf{X}$, dictionary $\mathbf{D}$, testing data $\mathbf{X_{test}}$, class label matrix $\mathbf{H}$ and parameters λ and τ

Step 1: Obtain the low-rank representation $\mathbf{Z}$ from $\mathbf{X}$ by optimizing (2)

Step 2: Perform low-rank decomposition on $\mathbf{X_{test}}$ with dictionary $\mathbf{D}$

$$\min_{\mathbf{Z}_{\text{test}},\mathbf{E}_{test}} \|\mathbf{Z}_{\text{test}}\|_* + \lambda\|\mathbf{E}_{test}\|_l \qquad s.\,t. \quad \mathbf{X}_{test} = \mathbf{DZ}_{test} + \mathbf{E}_{test},$$

Step 3: Given the low-rank representation $\mathbf{Z}$ of training data $\mathbf{X}$, learn the linear classifier $\hat{\mathbf{W}}$ by Eq. (3)

Output: For each z_i, identity $(z_i) \leftarrow \arg\max_k \quad (s = \hat{\mathbf{W}}z_i)$

$$k = \arg\max_k \quad (s = \hat{\mathbf{W}}z_i), \tag{4}$$

where s is the class label vector of test sample representation z_i. Algorithm 1 summarizes the procedure of integrating low-rank matrix recovery and linear classifier for recognition.

3. Learning robust and discriminative low-rank representation for face recognition

3.1. Motivation

LRR has been successfully employed to segment data drawn from a union of multiple linear (or affine) subspaces. It aims at finding the lowest-rank representation of a collection of vectors jointly. Also, LRR gives a way to recover the corrupted data drawn from multiple subspaces. We give an example by selecting a set of face images from the Extended Yale B database [71]. Each image has 96×84 pixels. For each subject, three images are randomly selected and imposed block

occlusion. In this case, for classification purpose, it is desired to obtain the robust and discriminative representation and remove the occlusion simultaneously. In Fig. 1, we decompose (a) into (b) and (c). The images in Fig. 0.1(b) (i.e., **DZ**) are called "corrected images" and the ones in Fig. 0.1(c) (i.e., **E**) are called "error images". By capturing the global structures and correcting the corruptions in a joint way, we have the following observations: (i) when the subspaces are independent, it is proved that there is a lowest-rank representation that reveals the membership of the samples: the within-cluster affinities are dense, and the between-cluster affinities are zeroes; (ii) each "error images" itself can be regarded as a low-rank matrix, refer to Fig. 0.1(c).

The block-diagonal affinity matrix **Z** can capture the global structure of data and encode the pairwise affinities between data vectors: samples within each class will demonstrate similar basic structures, while between-class samples will have different structures. The matrix **Z** can form a low-dimensional vector representation space and be assumed to be low-rank. Nevertheless, for "error images", we consider it individually in the original image space. Our purpose is to decompose the data matrix **X** into two parts **DZ** and **E**, where **Z** is a low-rank matrix in vector representation space, while **E** contains a series of low-rank noise images in original image space.

3.2. Problem formulation

Given a set of n images $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n \in \Re^{p\times q}$ from k classes, each image $\mathbf{X}_i$ may be contaminated by noise (occlusion, corruption, illumination changes, etc.). Our goal is to decompose each image as $\mathbf{X}_i = \mathbf{D}(\mathbf{Z}_i) + \mathbf{E}_i$, where $\mathbf{D}(\mathbf{Z}_i) = z_{1i}\mathbf{D}_1 + z_{2i}\mathbf{D}_2 + ... + z_{ti}\mathbf{D}_t$, and the model is formulated as:

$$\min_{Z,\,D,\mathbf{E}_i} \quad rank\,(\mathbf{Z}) + \lambda \sum_{i=1}^{n} rank\,(\mathbf{E}_i) \qquad s.\,t. \quad \mathbf{X} = \mathbf{DZ} + E, \tag{5}$$

where $\mathbf{X}=[\mathrm{vec}(\mathbf{X}_1), \mathrm{vec}(\mathbf{X}_2), ..., \mathrm{vec}(\mathbf{X}_n)]$, $\mathbf{D}=[\mathrm{vec}(\mathbf{D}_1), \mathrm{vec}(\mathbf{D}_2), ..., \mathrm{vec}(\mathbf{D}_t)]$ and $\mathbf{E}=[\mathrm{vec}(\mathbf{E}_1), \mathrm{vec}(\mathbf{E}_2), ..., \mathrm{vec}(\mathbf{E}_n)]$. The rank minimization problem is NP-hard, and in most cases there are no efficient algorithms that yield exact solutions. A popular heuristic is to replace the rank function with the nuclear norm. The nuclear norm is a convex relaxation of rank function and it has been shown that nuclear norm based models can obtain the optimal low rank solution in a variety of scenarios [61]. The following nuclear norm optimization problem provides a good surrogate for problem (5):

$$\min_{Z,\,D,\mathbf{E}_i} \quad \|\mathbf{Z}\|_* + \lambda \sum_{i=1}^{n} \|\mathbf{E}_i\|_* \qquad s.\,t. \quad \mathbf{X} = \mathbf{DZ} + \mathbf{E}. \tag{6}$$

The data matrix can be rewritten as $\mathbf{X} = [\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_k]$ ($\mathbf{Y}_i \in \Re^{m\times n_k}$ denotes the dataset of the ith class, each column of $\mathbf{Y}_i$ is a sample of class i and n_k denotes the training number of class k) and the dictionary $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_k]$ contains k sub-dictionaries where $\mathbf{D}_i$ corresponds to class i. The representation coefficients of $\mathbf{X}_i$over $\mathbf{D}$ can be denoted as $\mathbf{Z}_i = [\mathbf{Z}_{i,1}; \mathbf{Z}_{i,2}; ...; \mathbf{Z}_{i,k}]$, where $\mathbf{Z}_{i,j}$is the representation coefficient of $\mathbf{X}_i$ corresponds to $\mathbf{D}_j$. To obtain a low-rank and discriminative data representation, the dictionary **D** should have a powerful discriminative ability. The latent dictionary can encourage the data from the same class to have similar representations and those from different classes to have dissimilar representations. In other words, the sub-dictionary $\mathbf{D}_i$should be able to well represent $\mathbf{Y}_i$, and there is $\mathbf{Y}_i = \mathbf{D}_i\mathbf{Z}_{i,i} + \mathbf{E}_i$. $\mathbf{Z}_{i,j}$, the representation coefficients of $\mathbf{Y}_i$over $\mathbf{D}_j (i \neq j)$, are nearly all zeroes. It means that the non-zero coefficients of samples $\mathbf{Y}_i$ will only sparsely concentrate on the sub-dictionaries $\mathbf{D}_i$.

Suppose that $\mathbf{Q} = [q_1, q_2, ..., q_s] \in \Re^{t\times n}$ is the representation codes for **X** over dictionary **D**. We say that $\mathbf{Q} = q_i = [0,...,1,1,..., 0] \in \Re^t$ is an ideal discriminative representation codes for classification corresponding to the input signal x_i if the nonzero elements of q_i occur at those indices where the input signal $x_i$and the dictionary atom d_k share the same label. For example, assuming data matrix $\mathbf{X}=[x_1, x_2, ..., x_7]$ and dictionary $\mathbf{D}=[d_1, d_2, ..., d_6]$, where x_1, x_2, d_1 and d_2 are from class 1, x_3, x_4, x_5, d_3 and d_4 are from class 2, and x_6, x_7, d_5 and d_6 are from class 3, **Q** can be constructed as

$$\mathbf{Q} \equiv \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

where each column corresponds to the discriminative representation codes for an input signal. Although this decomposition might not lead to minimal reconstruction error, low-rank and sparse **Q** is an optimal discriminative representation for classification.

With the above definition and inspired by the work in [67, 70], we also incorporate the label consistency regularization term into the objective function. The regularization term encourages the desired representation **Z** to be close to the optimal discriminative representation **Q**. To obtain discriminative representation, the objective function in (6) is rewritten as

$$\min_{Z,\,D,\,\mathbf{E}_i} \quad \|\mathbf{Z}\|_* + \lambda \sum_{i=1}^{n} \|\mathbf{E}_i\|_* + \alpha\|\mathbf{Z} - \mathbf{Q}\|_F^2 \qquad s.\,t. \quad \mathbf{X} = \mathbf{DZ} + E, \tag{7}$$

where α control the contribution of regularization term.

The optimal representation of training and testing data can be directly used for classification by a multiclass linear classifier as described in Section 2.2. However, due to the learning of the representation and classifier are separated, the learned representation may not be optimal for final classification purpose. As in [70], we aim to include the classification error in the objective function to make the leaned representation optimal for classification. Our final objective function can be formulated as follows:

$$\min_{Z,D,W,\mathbf{E}_i} \quad \|\mathbf{Z}\|_* + \lambda \sum_{i=1}^{n} \|\mathbf{E}_i\|_* + \alpha\|\mathbf{Z} - \mathbf{Q}\|_F^2 \qquad s.\,t. \quad \mathbf{X} = \mathbf{DZ} + E,\ \mathrm{H} = \mathrm{WZ}, \tag{8}$$

where $\mathbf{W} \in \Re^{k\times t}$ denotes the classifier parameters. $\mathbf{H}=[h_1, h_2, ..., h_n] \in \Re^{k\times n}$ are the class label of input signals **X**. $h_i=[0, 0, ...,1, ...,0,0]^t \in \Re^k$ is the label vector of signal x_i, where the position of element "1" indicates the class of x_i. In model (8), term $\sum_{i=1}^{s} \|\mathbf{E}_i\|_*$ characterizes the individual reconstruction error, making the model robust to block occlusion; term $\|Z-Q\|_F^2$ denotes the discriminative representation error, making the representation codes discriminative between classes.

Compared with model (2), our model (8) uses a similar way to characterize the desired representation: samples within each class have similar structures, while between-class samples have different structures. The main difference between the two models is that model (2) converts all error images into vectors and stacks them to form an error matrix, which is assumed to be sparse. However, our model (8) looks at noise images in the original image space and assumes each of them to have low rank, characterized by nuclear norm.

There are some merits of using nuclear norm for structure error characterization:

(1) Nuclear norm can better reflect variations of structure error than L_1 or L_2 norm. Fig. 2 gives an example. In Fig. 2, image (a) is occluded by a baboon block as shown in (b). The error image between (b) and (a) is shown in (c). We re-arrange pixels of image (c) and obtain image (d). In previous methods, L_1 or L_2 norm are usually used to characterize the error image. However, these norms are based on pixel values, thus ignore the structural information of the error image. For example, the L_2 norm (or L_1 norm) values of image (c) is the same with that of image (d). Hence, it is difficult to distinguish the differences between (c) and (d). Fortunately, the nuclear norm values of images (c) and (d) are 47.75 and 58.14, respectively.
(2) From the distribution perspective, we can see that the distribution

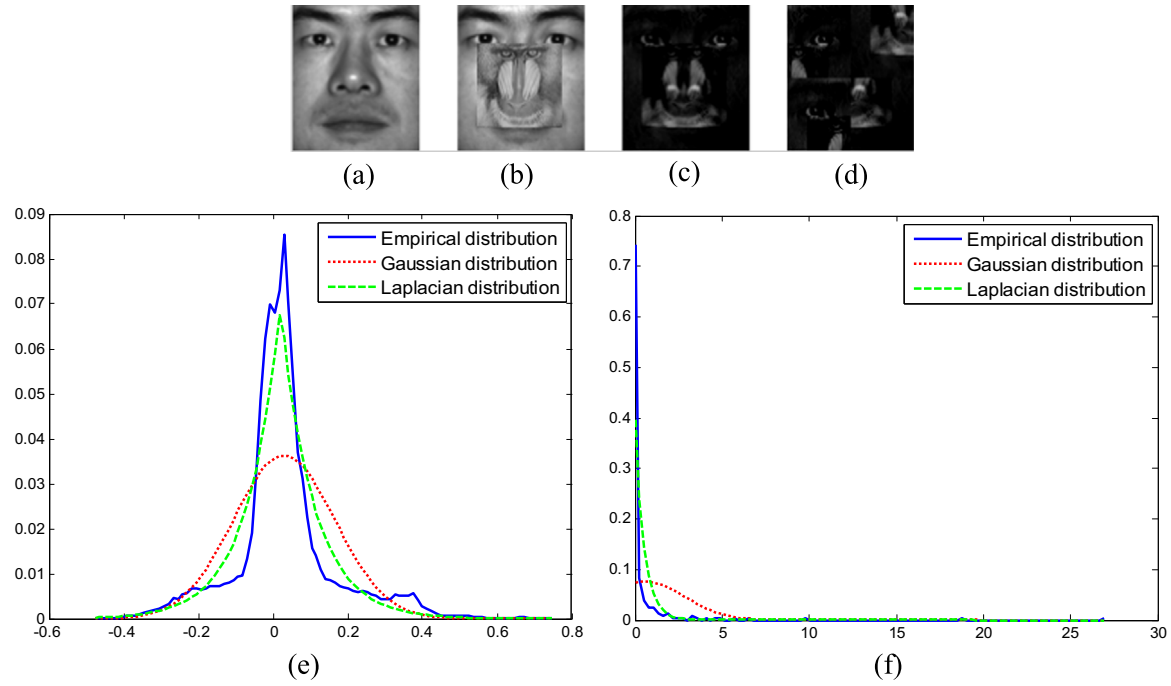

Fig. 2. (a) Original image; (b) observed image; (c) error image; (d) rearranged error image; (e) distributions of error image; and (f) distributions of singular values of error image.

of the error image does not follow Laplacian or Gaussian distribution in Fig. 2(e). As we all know, L_1 norm provides an optimal characterization for errors following the Laplacian distribution, while L_2 norm is optimal for Gaussian distribution. So, L_1 and L_2 norm cannot characterize this kind of structure error effectively. Fortunately, as it can be seen from Fig. 2(f) that the singular values of error image (c) fit the Laplacian distribution well. We know that nuclear norm is the sum of all singular values of a matrix, which can also be considered as L_1 norm of the singular value vector. Based on the above observations, we believe that nuclear norm is more suitable to describe the structural error than L_1 norm and L_2 norm.

3.3. Optimization via inexact ALM

Inexact Augmented Lagrange multiplier (ALM), also called the alternating direction method, has been used to solve many low-rank problem [61,72]. In this sub-section, we will detail how the ALM is.

adopted to solve problem (8) efficiently. To solve optimization problem (8), we first convert it to the following equivalent problem:

$$\min_{Z,D,J,W,\mathbf{E}_i} \quad \|\mathbf{J}\|_* + \lambda\sum_{i=1}^{n}\|\mathbf{E}_i\|_* + \alpha\|\mathbf{J}-\mathbf{Q}\|_F^2 \quad s.t. \quad \mathbf{X}=\mathbf{DZ}+E,\ \mathbf{H}=\mathbf{WZ},\ Z=J. \tag{9}$$

The augmented Lagrange function L is given by:

$$\begin{aligned} &L_\mu(\mathbf{Z},\ \mathbf{D},\mathbf{J},\mathbf{W},\mathbf{E}_i,\mathbf{Y}_1,\mathbf{Y}_2,\mathbf{Y}_3)\\ &=\|\mathbf{J}\|_* + \lambda\textstyle\sum_{i=1}^{n}\|\mathbf{E}_i\|_* + \alpha\|\mathbf{J}-\mathbf{Q}\|_F^2\\ &\quad + <\mathbf{Y}_1,\ \mathbf{X}-\mathbf{DZ}-\mathbf{E}> + <\mathbf{Y}_2,\ \mathbf{H}-\mathbf{WZ}> + <\mathbf{Y}_3,\ \mathbf{Z}-\mathbf{J}>\\ &\quad + \tfrac{\mu}{2}(\|\mathbf{X}-\mathbf{DZ}-\mathbf{E}\|_F^2 + \|\mathbf{H}-\mathbf{WZ}\|_F^2 + \|\mathbf{Z}-\mathbf{J}\|_F^2), \end{aligned} \tag{10}$$

where $<\mathbf{A},\mathbf{B}>=trace\ (\mathbf{A}^T\mathbf{B})$, $\mathbf{Y}_1$,$\mathbf{Y}_2$ and $\mathbf{Y}_3$ are the Lagrange multiplier, $\mu>0$ is a penalty parameter. The above problem is unconstrained. So it can be minimized with respect to **Z**, **D**, **J**, **W** and **E**, respectively, by fixing other variables, and then updating the Lagrange multiplier $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{Y}_3$.

For convenience, let us rewrite the augmented Lagrange function (10) as

$$\begin{aligned} &L_\mu(\mathbf{Z},\ \mathbf{D},\mathbf{J},\mathbf{W},\mathbf{E}_i,\mathbf{Y}_1,\mathbf{Y}_2,\mathbf{Y}_3)\\ &=\|\mathbf{J}\|_* + \lambda\textstyle\sum_{i=1}^{n}\|\mathbf{E}_i\|_* + \alpha\|\mathbf{J}-\mathbf{Q}\|_F^2\\ &\quad + \tfrac{\mu}{2}(\|\mathbf{X}-\mathbf{DZ}-E+\mathbf{Y}_1/\mu\|_F^2 + \|\mathbf{H}-\mathrm{WZ}+\mathbf{Y}_2/\mu\|_F^2\\ &\quad + \|\mathbf{Z}-J+\mathbf{Y}_3/\mu\|_F^2)\\ &\quad - \tfrac{1}{2\mu}(\|\mathbf{Y}_1\|_F^2 + \|\mathbf{Y}_2\|_F^2 + \|\mathbf{Y}_3\|_F^2), \end{aligned} \tag{11}$$

Updating J

Given **Z, D, W** and $\mathbf{E}_i$, the objective function L_μ in Eq. (11) can be rewritten as

$$\begin{aligned} &L_\mu(\mathbf{J}) = \|\mathbf{J}\|_* + \alpha\|\mathbf{J}-\mathbf{Q}\|_F^2 + \tfrac{\mu}{2}\|\mathbf{Z}-J+\mathbf{Y}_3/\mu\|_F^2\\ &=\|\mathbf{J}\|_* + \alpha Tr((\mathbf{J}-\mathbf{Q})^T(\mathbf{J}-\mathbf{Q}))\\ &\quad + \tfrac{\mu}{2}Tr((\mathbf{Z}-J+\mathbf{Y}_2/\mu)^T(\mathbf{Z}-J+\mathbf{Y}_3/\mu))\\ &=\|\mathbf{J}\|_* + \tfrac{2\alpha+\mu}{2}Tr\left(\mathbf{J}^T\mathbf{J}-2\left(\tfrac{2\alpha}{2\alpha+\mu}\mathbf{Q}^T+\tfrac{1}{2\alpha+\mu}\mathbf{Y}_3^T+\tfrac{\mu}{2\alpha+\mu}\mathbf{Z}^T\right)\mathbf{J}\right)\\ &\quad + const1\\ &=\|\mathbf{J}\|_* + \tfrac{2\alpha+\mu}{2}\left\|\mathbf{J}-\left(\tfrac{2\alpha}{2\alpha+\mu}\mathbf{Q}+\tfrac{1}{2\alpha+\mu}\mathbf{Y}_3+\tfrac{\mu}{2\alpha+\mu}\mathbf{Z}\right)\right\|F^2 + const2, \end{aligned} \tag{12}$$

where *const*1and *const*2 are constant terms, which are independent with the variable **J**. The optimization problem can be expressed as

$$\mathbf{J}^{k+1} = \arg\min_{\mathbf{J}} \frac{1}{2\alpha+\mu}\|\mathbf{J}\|_* + \frac{1}{2}\left\|\mathbf{J}-\left(\frac{2\alpha}{2\alpha+\mu}\mathbf{Q}+\frac{1}{2\alpha+\mu}\mathbf{Y}_3+\frac{\mu}{2\alpha+\mu}\mathbf{Z}\right)\right\|_F^2. \tag{13}$$

The optimal solution can be obtained by the singular value thresholding operator. Specifically, given a matrix $\mathbf{T}\in\Re^{P\times q}$ of rank r, its singular value decomposition (SVD) is

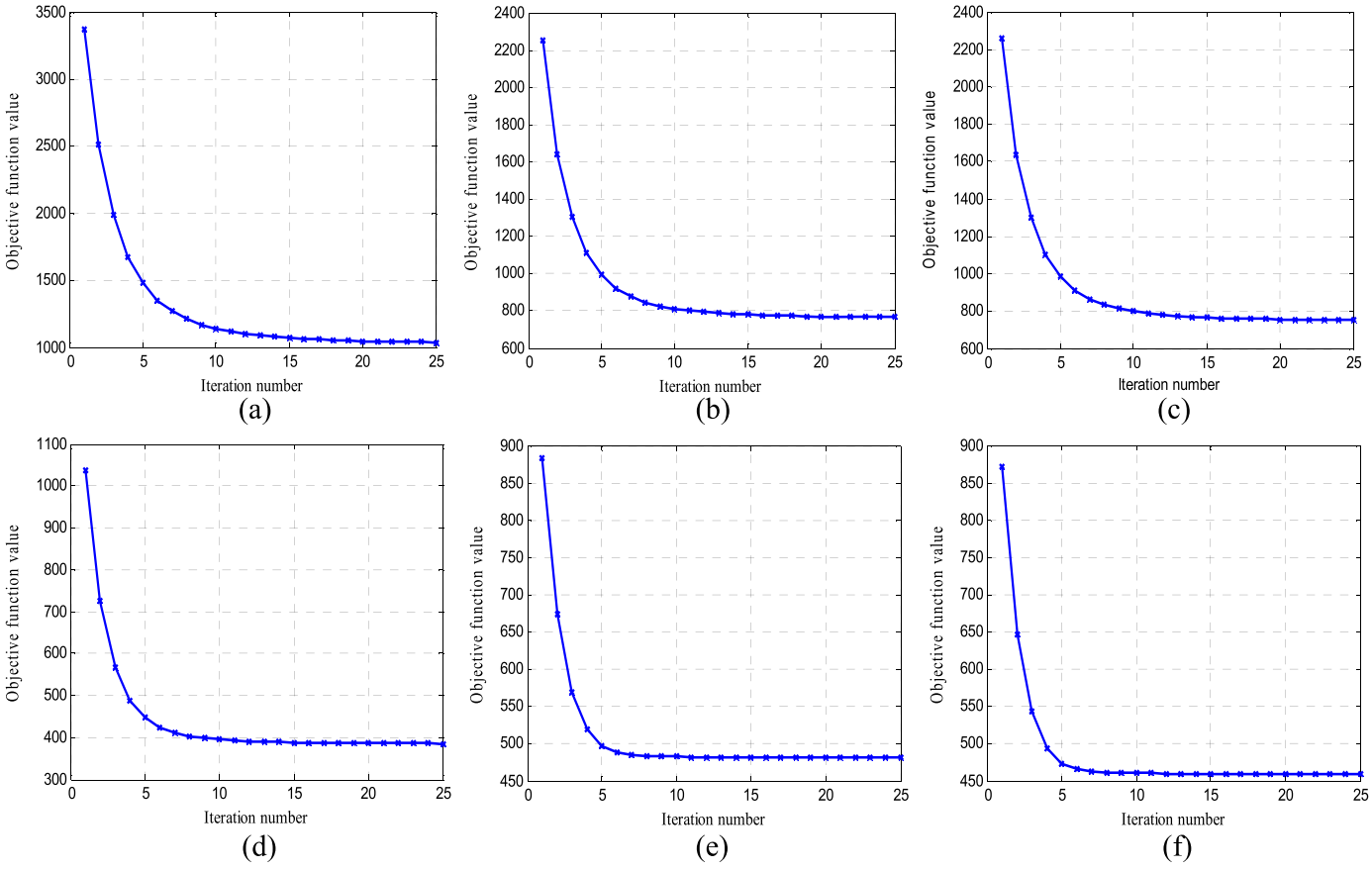

Fig. 3. Convergence analysis of RDLRR algorithm with face recognition on (a) Extended Yale B database with illumination; (b)–(c) AR database with sunglass and scarf; (d)–(f) Extended Yale B database occluded with unrelated block image, random block and mixed noise.

$$\mathbf{T} = \mathbf{U}_{p\times r}\mathbf{\Sigma}\mathbf{V}_{q\times r}^{T}, \quad \mathbf{\Sigma} = diag(\sigma_1,\ldots,\sigma_r), \tag{14}$$

Fig. 4. Samples images of a person under different illumination conditions in the Extended Yale B database from different sessions.

where σ_1, ..., σ_r are singular values, $\mathbf{U}$ and $\mathbf{V}$ are corresponding matrices with orthogonal columns. For a given $\tau > 0$, the singular value shrinkage operator $D_\tau(\bullet)$ is defined as follows

$$D_\tau(T) = \mathbf{U}_{p\times r} diag\left(\{\max(0,\ \sigma_j - \tau)\}_{1\le j\le r}\right)\mathbf{V}_{q\times r}^{T}. \tag{15}$$

Theorem 1. [73]: For matrix $\mathbf{T} \in \Re^{P\times Q}$ and $\tau > 0$, the singular value shrinkage operator in (15) obeys.

$$D_\tau(T) = \arg\min_{\mathbf{J}}\left(\tau\|\mathbf{J}\|_* + \frac{1}{2}\|\mathbf{J} - \mathbf{T}\|_F^2\right). \tag{16}$$

From Theorem 1, the optimal solution of (13) is

$$\mathbf{J}^{k+1} = D_{\frac{1}{2\alpha+\mu}}\left(\frac{2\alpha}{2\alpha+\mu}\mathbf{Q} + \frac{1}{2\alpha+\mu}\mathbf{Y}_3 + \frac{\mu}{2\alpha+\mu}\mathbf{Z}\right). \tag{17}$$

Updating Z.

Given $\mathbf{J}$, $\mathbf{D}$,$\mathbf{W}$ and $\mathbf{E}_i$, the optimization problem can be reformulated as

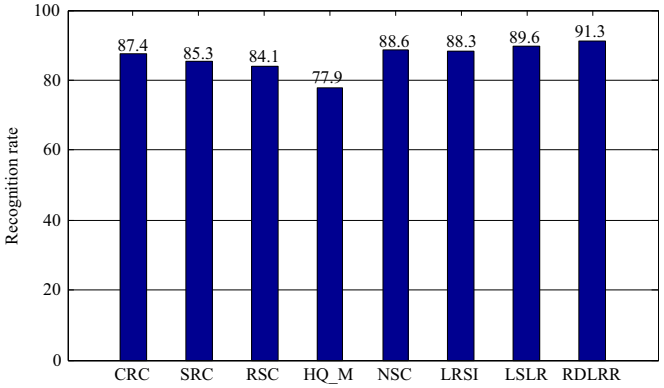

Fig. 5. Recognition rate (%) of each method under different illumination conditions on the Extended Yale B database.

$$\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}}\frac{\mu}{2}(\|\mathbf{X} - \mathbf{DZ} - E + \mathbf{Y}_1/\mu\|_F^2 + \|\mathbf{H} - \mathbf{WZ} + \mathbf{Y}_2/\mu\|_F^2 + \|\mathbf{Z} - J + \mathbf{Y}_3/\mu\|_F^2). \tag{18}$$

This equation is a quadratic form in variable $\mathbf{Z}$. Differentiating $L_\mu(\mathbf{Z})$ with respect to $\mathbf{Z}$, and let it be zero, we can obtain the optimal solution as follows:

$$\mathbf{Z}^{k+1} = (\mathbf{W}^T\mathbf{W} + \mathbf{D}^T\mathbf{D} + I)^{-1}\left(\mathbf{W}^T\mathbf{H} + \mathbf{D}^T\mathbf{X} - \mathbf{D}^T\mathbf{E} + \mathbf{J} + \frac{\mathbf{D}^T\mathbf{Y}_1 + \mathbf{W}^T\mathbf{Y}_2 - \mathbf{Y}_3}{\mu}\right). \tag{19}$$

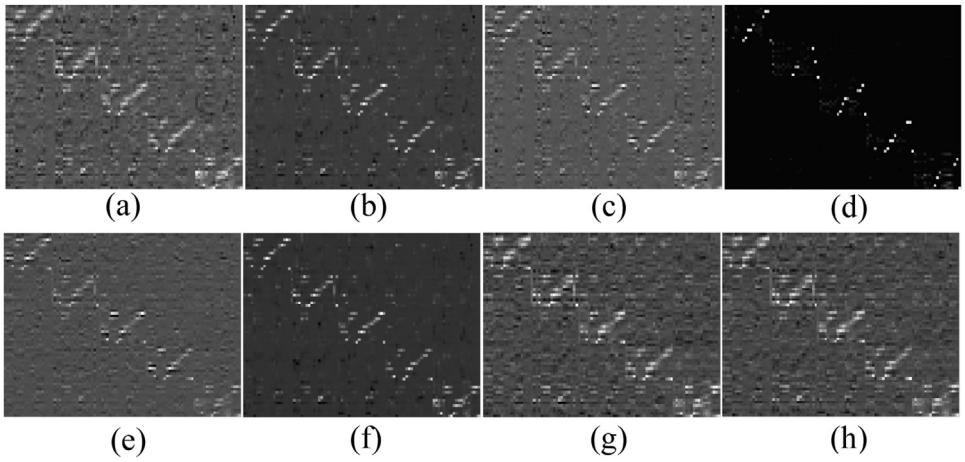

Fig. 6. Comparison of representations for testing samples from the first five classes on the Extended Yale B database. (a) CRC; (b) SRC; (c) RSC; (d) HQ_M; (e) NSC; (f) LRSI; (g) LSLR; (h) RDLRR.

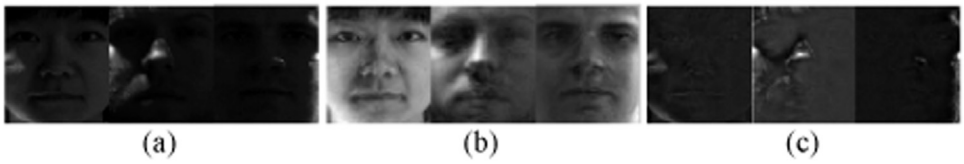

Fig. 7. Examples of image decomposition for testing samples on the Extended Yale B database. (a) original faces; (b) the recovered component DZ; (c) the noise component E.

Fig. 8. Samples images of a person under different illumination conditions in the CMU Multi-PIE database from different sessions.

Table 1
Comparisons of recognition rates (%) on the Multi-PIE database under different illuminations.

Methods	Session 2	Session 3	Session 4
CRC[36]	76.1	85.4	90.7
SRC[33]	77.8	86.2	92.6
RSC[49]	76.5	86.0	91.7
HQ_M[50]	77.8	87.9	91.2
NSC[52]	77.3	88.3	92.9
LRSI[59]	78.6	88.2	93.5
LSLR[67]	78.9	89.3	93.2
RDLRR	**80.3**	**91.2**	**94.8**

Updating D.

Given **J**, **Z**,**W** and $\mathbf{E}_{\boldsymbol{i}}$, **D** is the only variable in this sub-problem. So Eq. (11) can be rewritten as

$$\mathbf{D}^{k+1} = \arg\min_{\mathbf{D}} \frac{\mu}{2}(\|\mathbf{X} - \mathbf{DZ} - E + \mathbf{Y}_1/\mu\|_F^2). \tag{20}$$

Setting the partial derivative of L with respect to D equal to zero, we obtain

$$\mathbf{D}^{k+1} = (\mathbf{X} - E + \mathbf{Y}_1/\mu)\mathbf{Z}^T(\mathbf{ZZ}^T)^{-1}. \tag{21}$$

Updating E.

Given **Z, D, W** and **J**, the objective function L_{μ} in Eq. (11) can be rewritten as

$$\begin{aligned} L_{\mu}(\mathbf{E}) &= \lambda\textstyle\sum_{i=1}^{n}\|\mathbf{E}_i\|_* - <\mathbf{Y}_1,\ \mathbf{E}> + \frac{\mu}{2}\|X-\mathrm{DZ}-E\|_F^2 \\ &= \lambda\textstyle\sum_{i=1}^{n}\|\mathbf{E}_i\|_* - \sum_{i=1}^{n} sum(\mathbf{Y}_1^i\cdot\mathbf{E}_i) + \frac{\mu}{2}\sum_{i=1}^{n}\|\mathbf{X}_i-D(\mathbf{Z}_i)-\mathbf{E}_i\|_F^2 \\ &= \textstyle\sum_{i=1}^{n}\{\lambda\|\mathbf{E}_i\|_* - sum(\mathbf{Y}_1^i\cdot\mathbf{E}_i) + \frac{\mu}{2}\|\mathbf{X}_i-D(\mathbf{Z}_i)-\mathbf{E}_i\|_F^2\} \\ &= \textstyle\sum_{i=1}^{n}\left\{\lambda\|\mathbf{E}_i\|_* + \frac{\mu}{2}\left\|\mathbf{X}_i-D(\mathbf{Z}_i)-\mathbf{E}_i + \frac{1}{\mu}\mathbf{Y}_1^i\right\|_F^2\right\}, \end{aligned} \tag{22}$$

where $\mathbf{D}(\mathbf{Z}_i) = z_{1i}\mathbf{D}_1 + z_{2i}\mathbf{D}_2 + \ldots + z_{ti}\mathbf{D}_t$, each $\mathbf{D}_j(j=1,\ldots,t)$ is a matrix, $sum(\mathbf{X})$ sums all elements of matrix **X**, and symbol • denotes Hadamard product. Each $\mathbf{E}_i(i=1,\ldots,n)$ in (22) is separable and can be solved one by one. Thus, the optimization problem can be reformulated as

$$\mathbf{E}_i^{k+1} = \arg\min_{\mathbf{E}_i} \frac{\lambda}{\mu}\|\mathbf{E}_i\|_* + \frac{1}{2}\left\|\mathbf{E}_i - \left(\mathbf{X}_i - D(\mathbf{Z}_i) + \frac{1}{\mu}\mathbf{Y}_1^i\right)\right\|_F^2. \tag{23}$$

By Theorem 1, the optimal $\mathbf{E}_i^{k+1}$ can be obtained as follows:

$$\mathbf{E}_i^{k+1} = D_{\frac{\lambda}{\mu}}\left(\mathbf{X}_i - D(\mathbf{Z}_i) + \frac{1}{\mu}\mathbf{Y}_1^i\right). \tag{24}$$

Updating W.

Given **Z,J** and **E**, the optimization problem can be rewritten as

$$\mathbf{W}^{k+1} = \arg\min_{\mathbf{W}} \|\mathbf{H} - \mathrm{WZ} + \mathbf{Y}_2/\mu\|_F^2 + \eta\|\mathbf{W}\|_F^2. \tag{25}$$

We add a regularization term $\eta\|\mathbf{W}\|_F^2$ into (25) to make the solution

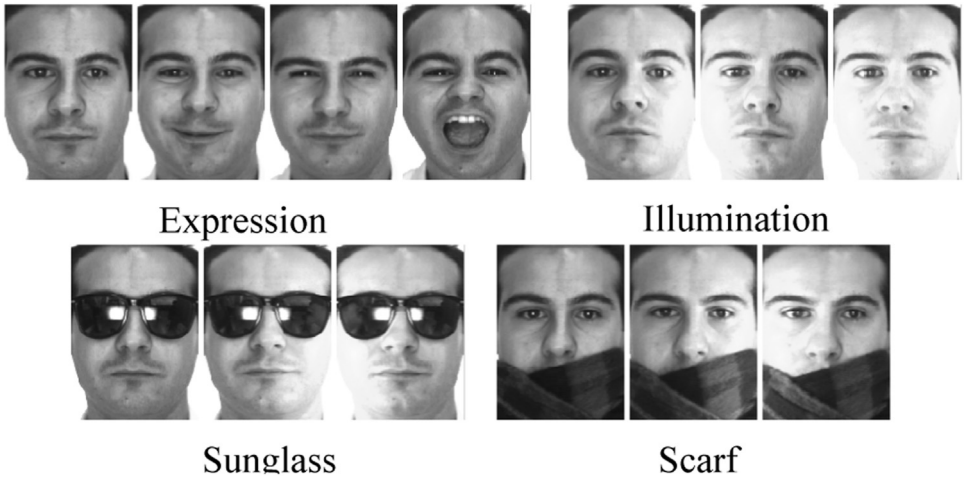

Fig. 9. Sample images of a person from Session 1 in the AR database.

Table 2
Comparisons of recognition rates (%) with different percentages of disguise occlusion in the training set.

Methods	3/(2+3)=60%		3/(3+3)=50%		3/(4+3)=43%		3/(5+3)=38%	
	Sunglasses	Scarf	Sunglasses	Scarf	Sunglasses	Scarf	Sunglasses	Scarf
CRC[36]	87.3	80.7	88.3	80.7	88.0	80.7	87.0	84.3
SRC[33]	88.7	80.0	90.0	81.3	89.3	82.7	88.0	85.3
RSC[49]	89.0	80.0	90.3	81.3	90.3	85.9	88.7	84.0
HQ_M[50]	89.3	76.7	89.3	79.7	89.7	78.0	88.0	83.3
NSC[52]	89.7	83.7	89.7	84.0	89.7	86.0	89.0	87.7
LRSI[59]	88.6	82.5	90.3	83.5	90.5	84.3	89.0	86.4
LSLR[67]	89.3	85.3	90.3	86.7	90.7	86.2	90.0	87.0
RDLRR	**91.3**	**88.3**	**92.0**	**91.0**	**92.0**	**88.7**	**91.7**	**90.0**

Table 3
Comparisons of recognition rates (%) with different percentages of disguise occlusion in the training set.

Methods	Sunglasses+Scarf		
	2/(4+2)=33%	4/(2+4)=67%	6/(0+6)=100%
CRC[36]	77.3	78.3	79.0
SRC[33]	78.7	79.5	79.8
RSC[49]	79.2	79.3	80.7
HQ_M[50]	79.5	79.5	80.7
NSC[52]	79.8	80.3	81.0
LRSI[59]	79.3	80.0	80.5
LSLR[67]	80.0	80.1	80.5
RDLRR	**81.2**	**82.2**	**82.5**

more stable, where η is a small scalar variable (it is set as 0.001 in all our experiments). Since problem (25) is a standard Ridge regression model, we can get its close-form solution

$$\mathbf{W}^{k+1} = (\mathrm{H} + \mathbf{Y}_2/\mu)\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T + \eta\mathbf{I})^{-1}. \tag{26}$$

Algorithm 2. Solving Problem (8) by Inexact ALM.

Input: A set of Data matrices $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n \in \Re^{p\times q}$, Dictionary $\mathbf{D}$, Parameters λ, α,
Output: $\mathbf{Z}$, $\mathbf{E}$
Initialize: $\mathbf{W}^o, \mathbf{Z}^o = \mathbf{J}^o = \mathbf{E}^o = \mathbf{Y}_1{}^o = \mathbf{Y}_2{}^o = 0$,
$\mu = 10^{-6}, max_\mu = 10^1{}^o$, $\rho = 1.1$, $\varepsilon = 10^{-8}$
while $\|\mathbf{X} - \mathbf{D}^{k+1}\mathbf{Z}^{k+1} - \mathbf{E}^{k+1}\|_\infty > \varepsilon$ or $\|\mathbf{Z}^{k+1} - \mathbf{J}^{k+1}\|_\infty > \varepsilon$ **do**

Step 1: $\mathbf{J}^{k+1} = D_{\frac{1}{2\alpha+\mu}}\left(\frac{2\alpha}{2\alpha+\mu}\mathbf{Q} + \frac{1}{2\alpha+\mu}\mathbf{Y}_3 + \frac{\mu}{2\alpha+\mu}\mathbf{Z}\right)$.

Step 2: $\mathbf{Z}^{k+1} = (\mathbf{W}^T\mathbf{W} + \mathbf{D}^T\mathbf{D} + I)^{-1}\left(\mathbf{W}^T\mathbf{H} + \mathbf{D}^T\mathbf{X} - \mathbf{D}^T\mathbf{E} + \mathbf{J} + \frac{\mathbf{D}^T\mathbf{Y}_1 + \mathbf{W}^T\mathbf{Y}_2 - \mathbf{Y}_3}{\mu}\right)$.

Step 3: $\mathbf{E}_i^{k+1} = D_{\frac{\lambda}{\mu}}\left(\mathbf{X}_i - D(\mathbf{Z}_i) + \frac{1}{\mu}\mathbf{Y}_1^i\right)$.

Step 4: $\mathbf{D}^{k+1} = (\mathbf{X} - E + \mathbf{Y}_1/\mu)\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1}$

Step 5: $\mathbf{W}^{k+1} = (\mathrm{H} + \mathbf{Y}_2/\mu)\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T + \eta\mathbf{I})^{-1}$.

$\mathbf{Y}_1{}^{k+1} = \mathbf{Y}_1{}^k + \mu(\mathbf{X} - \mathbf{D}^{k+1}\mathbf{Z}^{k+1} - \mathbf{E}^{k+1})$

Step 6: $\mathbf{Y}_2{}^{k+1} = \mathbf{Y}_2{}^k + \mu(\mathbf{H} - \mathbf{W}^{k+1}\mathbf{Z}^{k+1})$

$\mathbf{Y}_3{}^{k+1} = \mathbf{Y}_3{}^k + \mu(\mathbf{Z}^{k+1} - \mathbf{J}^{k+1})$.

μ=min ($\rho\mu$, max_μ).
end while

Initialization

We need to initialize $\mathbf{W}^0$ for RDLRR. Given the initialized $\mathbf{D}^0$ ($\boldsymbol{X}$ in this paper), we apply the original low-rank model (i.e., model (2)) to compute the low-rank representation $\mathbf{Z}$ for training signals $\mathbf{X}$. To initialize $\mathbf{W}^0$, we employ the multivariate ridge regression model, with the quadratic loss and L_2 norm regularization, as follows:

$$\mathbf{W} = \arg\min_{\mathbf{W}} \|\mathbf{H} - \mathbf{W}\mathbf{Z}\|_F^2 + \eta\|\mathbf{W}\|_F^2, \tag{27}$$

which yields the following solution:

$$\mathbf{W} = \mathbf{H}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T + \eta\mathbf{I})^{-1}. \tag{28}$$

In summary, the pseudo code of our method to solve problem (8) is

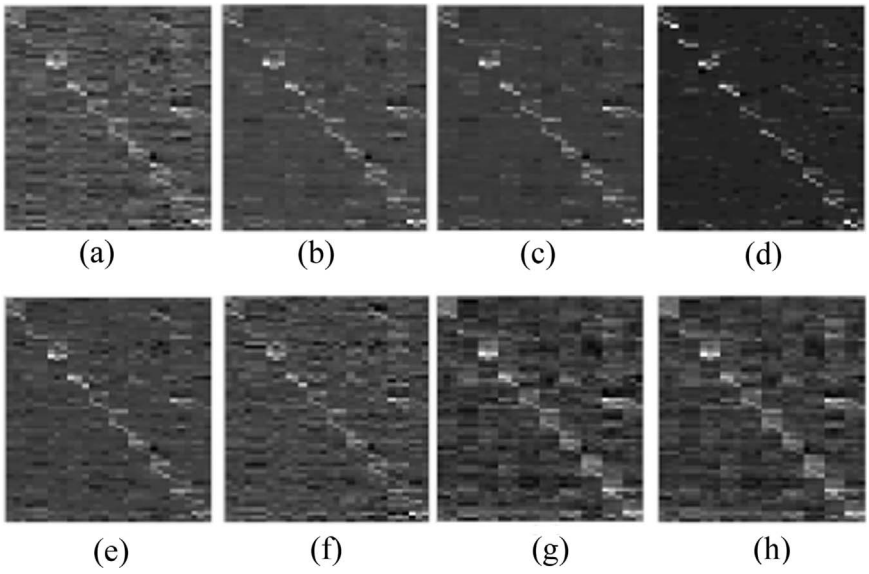

Fig. 10. Comparison of representations for testing samples from the first ten classes on the AR database. (a) CRC; (b) SRC; (c) RSC; (d) HQ_M; (e) NSC; (f) LRSI; (g) LSLR; (h) RDLRR.

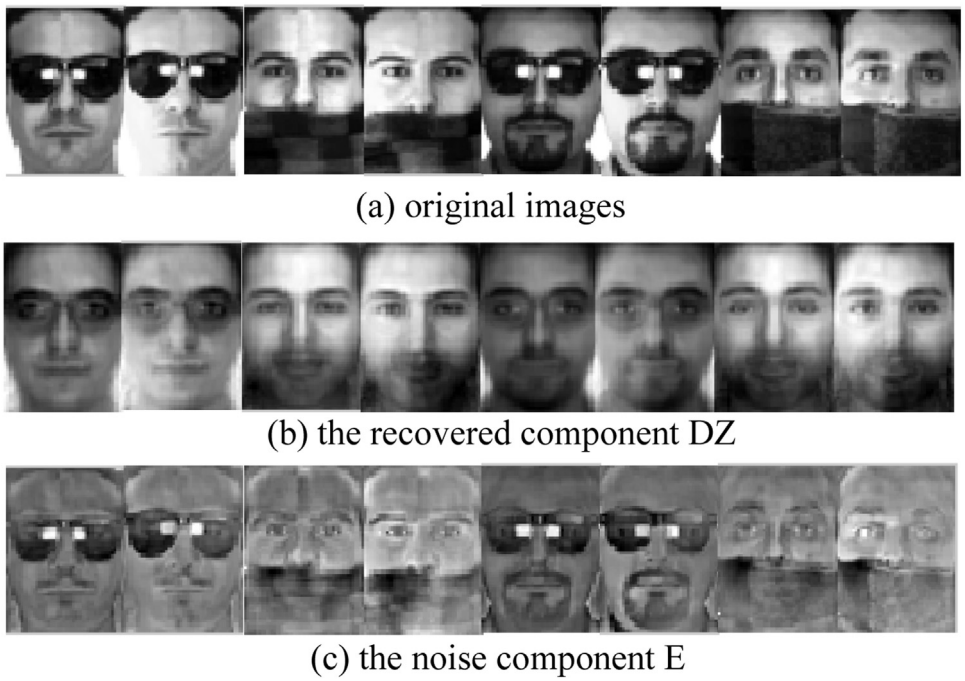

Fig. 11. Examples of image decomposition for testing samples from two classes on the AR database.

shown in Algorithm 2.

3.4. Complexity analysis

Given the training sample size n and the image size $p\times q$, let $m=p\times q$. For the convenience of analysis, we assume that the sizes of both $\mathbf{X}$ and $\mathbf{D}$ are $m\times n$ in the following. The major computation of Algorithm 2 is at Step 1 and Step 4, which requires computing the SVD of an $n\times n$ matrix and n $p\times q$ matrices. For a matrix in $\Re^{n\times n}$, the exact SVD has a computational complexity of $O(n^3)$, while for n matrices in $\Re^{p\times q}$, the computational complexity is $O(\min(p^2q, pq^2))$. In the case that $p=q$, the computational complexity becomes $O(m^{1.5})$. So, the computational complexity of the algorithm is $O(n^3+nm^{1.5})$.

When the number of data samples n is large, the SVD process in Step 1 will be time consuming. Fortunately, the computational cost can be reduced by Theorem 4.3 of [64], which concludes that the optimal solution $\mathbf{Z}^*$ (with respect to the variable $\mathbf{Z}$) to (9) always lies within the subspace spanned by the rows of $\mathbf{D}$, i.e., $\mathbf{Z}^*$ can be expressed as $\mathbf{Z}^* = \mathbf{P}^*\widetilde{\mathbf{Z}}^*$, where $\mathbf{P}^*$ can be computed in advance by orthogonalizing the rows of $\mathbf{D}$. Thus, problem (9) can be equivalently transformed into the following problem by replacing $\mathbf{Z}$ with $\mathbf{P}^*\widetilde{\mathbf{Z}}$:

$$\min_{\widetilde{\mathbf{Z}},A,\mathbf{B},\mathbf{E}_i} \quad \|\widetilde{\mathbf{Z}}\|_* + \lambda\sum_{i=1}^{n}\|\mathbf{E}_i\|_* + \alpha\|\mathbf{P}^*\widetilde{\mathbf{Z}} - \mathbf{Q}\|_F^2$$
$$s.\,t. \quad \mathbf{X} = A\widetilde{\mathbf{Z}} + E,\ \mathrm{H} = \mathbf{B}\widetilde{\mathbf{Z}},$$

where $\mathbf{A}=\mathbf{D}\mathbf{P}^*$, $\mathbf{B}=\mathbf{W}\mathbf{P}^*$. Since the number of columns of $\mathbf{A}$ is at most m (assuming $m\le n$), the above problem can be solved with a complexity of

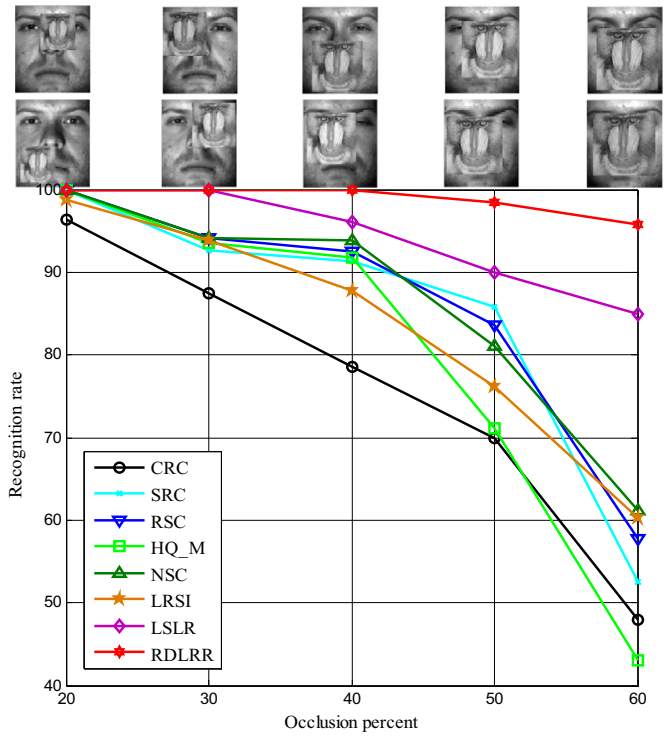

Fig. 12. The recognition rates (%) of CRC, SRC, RSC, HQ_M, NSC, LRSI, LSLR and RDLRR with the occlusion (unrelated block image) percentage ranging from 20 to 60.

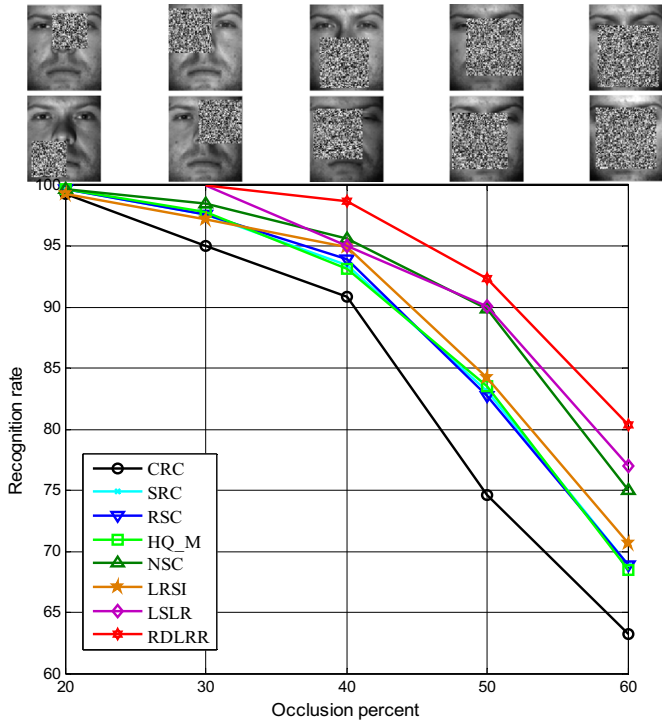

Fig. 13. The recognition rates (%) of CRC, SRC, RSC, HQ_M, NSC, LRSI, LSLR and RDLRR with the occlusion (square random block) percentage ranging from 20 to 60.

$O(m^2n+ nm^{1.5})$ by a similar way as Algorithm 2.

Considering the cost of orthogonalization and the number of iterations, the complexity of Algorithm 2 is $O(m^2n)$ $+O(k(m^2n+ nm^{1.5}))$, where k is the number of iterations. The iteration number k mainly depends on the choice of ρ: k is smaller while ρ is larger, and

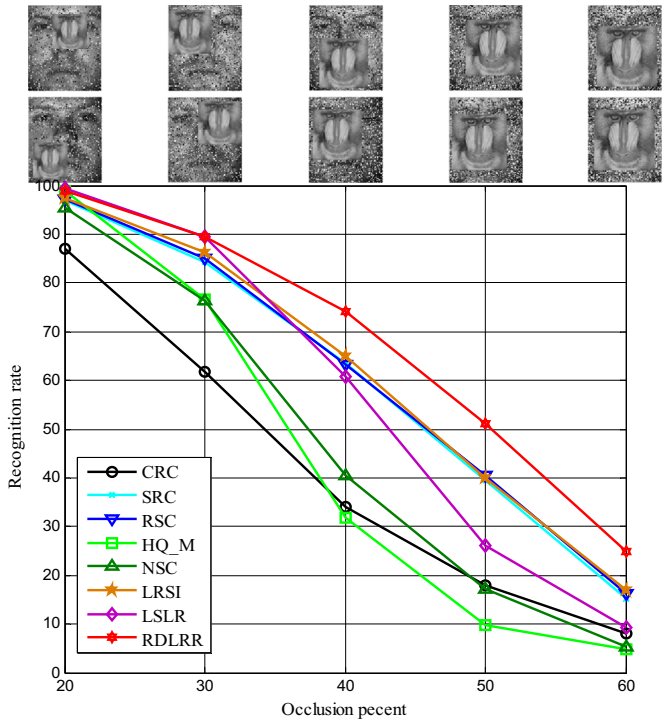

Fig. 14. The recognition rates (%) of CRC, SRC, RSC, HQ_M, NSC, LRSI, LSLR and RDLRR with the occlusion (mixture noise) percentage ranging from 20 to 60.

vice versa. Although larger ρ may speed up the algorithm, it has the risk of losing optimality to use large ρ[72]. Empirically, we always set ρ=1.1.

3.5. Convergence analysis

For inexact ALM, which is a variation of exact ALM, its convergence has been well studied when the number of blocks is at most two [72]. Up to now, it is still difficult to generally ensure the convergence of inexact ALM with three or more blocks [74]. Since there are three blocks in Algorithm 2 and the objective function (8) is not smooth, it would be not easy to prove the convergence of Algorithm 2 in theory.

Fortunately, there actually exist some guarantees for ensuring the convergence of Algorithm 2. According to the theoretical results in [75], two conditions are sufficient (but may not necessary) for Algorithm 2 to

converge: The first condition is that the dictionary matrix $\mathbf{D}$ is of full column rank; the second one is that the optimality gap produced in each iteration step is monotonically decreasing, namely, the error

$$\varepsilon_k = \left\| (\mathbf{Z}^k, \mathbf{J}^k) - \arg\min_{\mathbf{Z},\mathbf{J}} L_\mu \right\|_F^2$$

is monotonically decreasing, where $\mathbf{Z}^k$ (respectively, $\mathbf{J}^k$) denotes the solution produced at the k^{th} iteration, $\arg\min_{\mathbf{Z},\mathbf{J}} L_\mu$ indicates the "ideal" solution obtained by minimizing the Lagrange function L_μ with respect to both $\mathbf{Z}$ and $\mathbf{J}$ simultaneously. The first condition is easy to obey since (8) can be converted into an equivalent problem where the full column rank condition is always satisfied (see Section 3.3). For the monotonically decreasing condition, although it is not easy to strictly prove it, the convexity of the Lagrange function could guarantee its validity to some extent [75].

In addition, we find that the RDLRR algorithm converges asymptotically in our experiments. Fig. 3 illustrates the convergence of

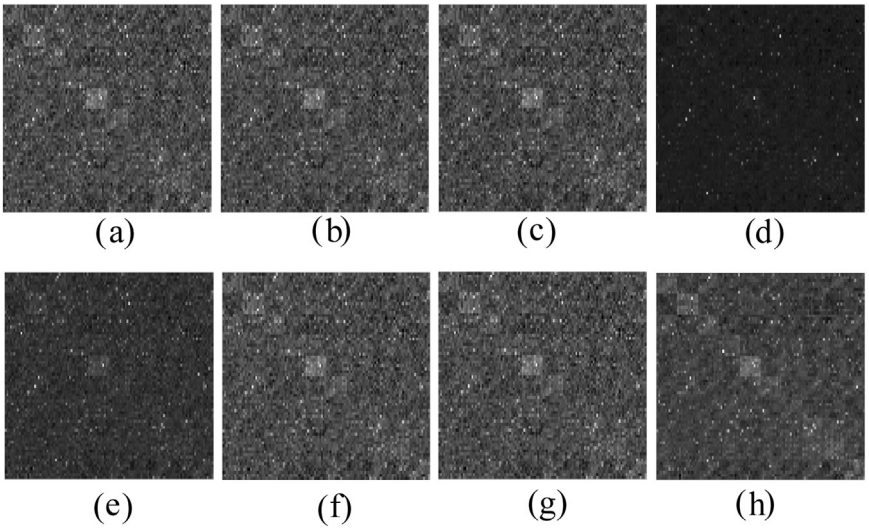

Fig. 15. Comparison of representations for testing samples from the first ten classes on the Extended Yale B database. (a) CRC; (b) SRC; (c) RSC; (d) HQ_M; (e) NSC; (f) LRSI; (g) LSLR; (h) RDLRR.

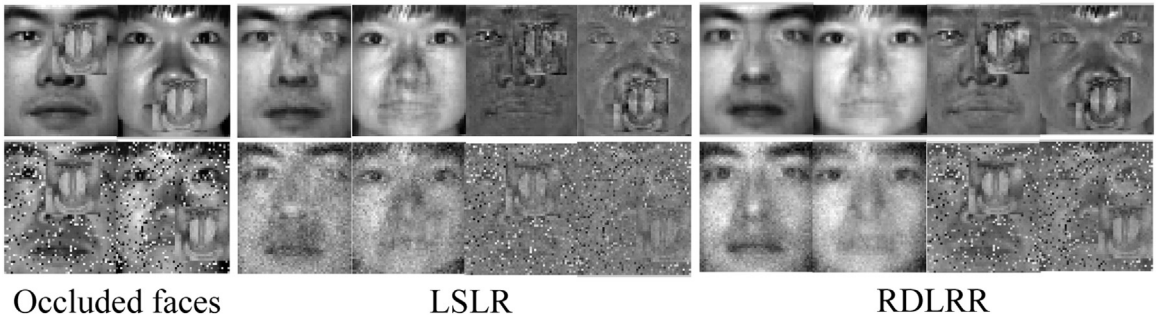

Fig. 16. Recovered images and occluded parts for testing samples on the Extended Yale B database.

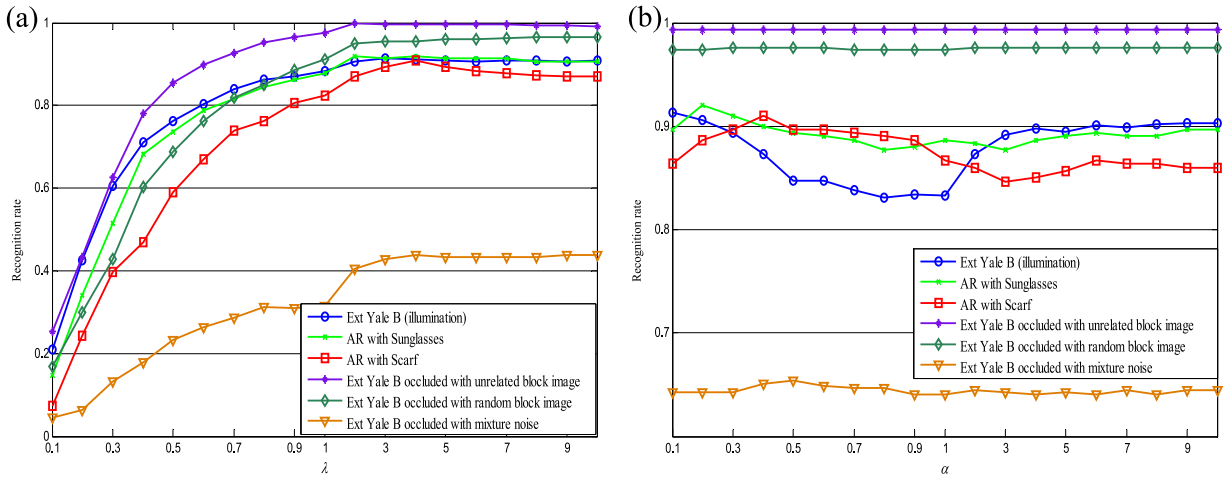

Fig. 17. Recognition rates of RDLRR with different parameters in different face recognition scenarios. (a) Parameter λ and (b) parameter α.

RDLRR on some face database. Fig. 3(a) demonstrates that the value of the objective function (8) almost tends to be stable after 15 iterations on Extended Yale B database with illumination. Fig. 3(b)–(c) shows that the value of the objective function (8) decreases fast and converges after 15 iterations on AR database with sunglass and scarf. Fig. 3(d)–(f) indicates that the value of the objective function (8) becomes stable after 10 iterations on Extended Yale B database occluded with unrelated block image, random block and mixed noise, respectively.

4. Experimental results and discussions

In this section, we evaluate our method on several face databases and compare it with state-of-the-art methods: CRC [36], SRC [33], RSC [49], HQ_M (Multiplicative form) [50], NSC [52], LRSI [59] and LSLR [67].

CRC and SRC are tuned to achieve their best performance by choosing the optimal regression parameters, while the parameter settings of other methods follow the authors' suggestions. It should be mentioned that here all experiments are done on the original face images, without any feature extraction and image preprocessing step. Also, in our experiments, both training and testing data can be occluded.

4.1. Recognition with varying illumination

In this section, we evaluate the proposed method under different illumination conditions. The first experiment was conducted on the Extended Yale B database, which contains 38 human subjects under 9 poses and 64 illumination conditions [71]. The 64 images of a person in a particular pose are acquired at a camera frame rate of 30 frames per second. So the variations in head poses and facial expressions are small for those 64 images. All frontal-face images marked with P00 are used in this experiment, and each is resized to 96×84=8064 pixels. The database is divided into five Subsets (see Fig. 4): Subset 1 is consisted of 266 images (seven images per subject) under normal lighting conditions; Subsets 2 and 3, each includes 12 images per subject, characterizes slight-to-moderate illumination variations, while Subset 4 (14 images per subject) and Subset 5 (19 images per subject) illustrate severe light variations. As we know, extreme illumination change is a challenging task for most face recognition methods. Therefore, we use Subset 4 for training and Subset 5 for testing.

Fig. 5 shows the recognition rates of all methods. From Fig. 5, we can see that the proposed RDLRR achieves the best results among all methods. Due to the presence of illumination in training set, the (robust) regression based methods like SRC, RSC and HQ_M seem not very robust to extreme illumination changes. By characterizing the structure error using nuclear norm, NSC can produce better performance than SRC, RSC and HQ_M. Both LRSI and LSLR obtain the similar results as NSC in this case. Fig. 6 illustrates the representations for the first five subjects. There are 14×5=70 training images and 19×5=95 testing images. The first line shows the testing images' representations based on CRC, SRC, RSC and HQ_M. Figs. 6(e), 6(f) and 6(g) are the representations based on NSC, LRSI and LSLR. In our learned representation, shown in Figs. 6(h), images from the same class show strong similarities. This representation is much more discriminative than the others. We also show some decomposition results in Fig. 7. The first three are original images. The middle and the last three images are the recovered component and noise component, respectively.

The CMU Multi-PIE database [76], which contains face images of 337 subjects recorded in four different sessions, was used in the second experiment. In our experiments, we consider the training set of all 249 subjects in Session 1. For each of the 249 subjects, we select 7 frontal neutral images with slight illumination changes for training. Thus, training set has a total of 7×249=1,743 images. All subjects of Session 2, 3 and 4, each having 10 frontal neutral images with different illumination variations are used or testing. All face images are manually cropped and resized to 50×40=2,000 pixels. Example images from the CMU Multi-PIE database are shown in Fig. 8. Table 1 tabulates recognition rates of all methods for the three test sets. From Table 1, we can see that the proposed RDLRR always achieves the best performance among all the methods. The robust representation methods like SRC, RSC, HQ_M and NSC also achieve similar results in these tests. Note that the illumination conditions of images in the Multi-PIE database are much better than those in the Extended Yale B database. Those methods (such as LRSI, LSLR) designed for recognition cases where both training and testing images are occluded achieve competitive results in these tests.

4.2. Recognition with real disguise

The AR database [77] contains over 4000 color face images for 126 people, including frontal views of faces with different facial expressions, lighting conditions and occlusions. For each subject, twenty-six face images are taken in two separate sessions. There are thirteen images for each session, in which three images with sunglasses, another three with scarfs, and the remaining seven are with illumination and expression changes and they are considered as clean/neutral images (see Fig. 9 for example). All images are downsampled to 55×40=2200 pixels and converted to gray scale. In this experiment, a subset containing 50 male and 50 female subjects was chosen. Experiments are conducted under three different scenarios:

Sunglasses: We first consider occluded training images with the presence of sunglasses, which occlude about 20% of the face. We randomly select n_c neutral image(s) plus 3 images with sunglasses from Session 1 for training, and 3 images with sunglasses from Session 2 for testing. To assess the influence of the ratio $3/(n_c+3)$ for robust face recognition, we vary the number of n_c from 2 up to 5.

Scarf: We consider occluded training images occluded by disguise due to the presence of Scarfs, which occlude about 40% of the face. The choice of the training and testing set is the same as that for the above (**Sunglasses**) case.

Sunglasses+Scarf: For this case, the training images are occluded due to the presence of sunglasses and scarf. From Session 1, we randomly select n_{sg} image(s) with sunglasses plus n_{sc} image(s) with scarf for training (we fix $n_{sg}+n_{sc}=6$). The numbers of n_{sg} and n_{sc} are set to be the same, and they range from 1 to 3. The testing set consists of 3 images with sunglasses and 3 images with scarf (all from Session 2). Note that the setting of this scenario is different from those in **Sunglasses** and **Scarf**. The number of training images in the above two cases varies with n_c, while the number of training images in this scenario is fixed at 6.

Tables 2, 3 show the recognition results of the above three scenarios using different approaches. From these two tables, we can see that our method generally outperforms all other methods across different settings. In Table 2, we observe that for testing images with sunglass, where the occlusion level (about 20%) is relatively low, there is no significant performance difference between all these methods. However, when the occlusion level becomes larger (about 40%), in the case of images with scarves, the advantage of the proposed RDLRR becomes evident. In addition, the difference in recognition rates of other methods between the two scenarios is large, which demonstrates that these methods are sensitive to the type of occlusions to some extent. In contrast, our method has much smaller performance gap, which illustrates that our method is much less sensitive to the type of occlusions in the training set.

We visualize the representations for the first ten subjects under the scarf scenario. There are 6×10=60 training images and 3×10=30 testing images. Figs. 10(a)–(d) shows the testing images' representations based on CRC, SRC, RSC and HQ_M. Figs. 10(e)–(h) are the representations based on NSC, LRSI, LSLR and our method. The testing images automatically generate a block diagonal structure in LSLR and our method, which is absent in other approaches.

The image decomposition examples are shown in Fig. 11. The first row is the original occluded images. The second shows the recovered component and the third is the separated noises. Our approach can remove real disguise such as sunglasses and scarves from the original images.

4.3. Recognition with random block occlusion

In this part, we test the robustness of our method to block occlusion. We use Subset 1 of Extended Yale B for training and Subset 2 for testing. In the first experiment, both the training and testing images are corrupted by a randomly located square block of "baboon" image with varying block size. The block size determines the occlusion level of an image. Fig. 12 plots the recognition rates of each method versus different occlusion levels (from 20% to 60%). We can

see that with the increase of the occlusion levels, the proposed RDLRR begins to outperform other methods. When occlusion level is no more than 30%, LSLR achieves similar performance with RDLRR. When the occlusion level becomes larger (more than 50%), the advantage of RDLRR becomes evident. When occlusion level is 60%, the recognition rate of RDLRR is 10%, 34.6%, 38.1% and 43.3% higher than that of LSLR, NSC, RSC and SRC, respectively.

The setting of the second experiment is similar to that of the first one. The only difference is that Subset 2 with square random block whose elements are random numbers between 0 and 255 is used for testing. The recognition rates of each method versus various occlusion levels (from 20% to 60%) are shown in Fig. 13. Generally, the results in Fig. 13 are consistent with those in Fig. 12. The proposed RDLRR always achieves the best performance in both occlusion cases. The performance of RSC and HQ_M are not good in this case. NSC achieves comparable results when the occlusion level is low than 50%. The performance difference between RDLRR and LSLR are not remarkable as that in Fig. 12 when the occlusion level is equal or larger than 50%. The recognition rates of CRC drop fast with the increase of the occlusion levels.

In the third experiment, we also use Subset 1 and Subset 2 for training and testing respectively. Each image is occluded with the mixture noise (pixel corruption and block occlusion). Fig. 14 plots the recognition rates of each method with different pixel corruption (and block occlusion) levels (from 20% to 60%). It can be seen from Fig. 14 that the performance of each method degrades with the increase of the mixture noise levels. However, the proposed RDLRR still achieves the best results among all the methods. The recognition rates of HQ_M and NSC are poor when facing the mixture noises. The probable reason may be that each of HQ_M and NSC is designed to address the single type noise.

Fig. 15 also shows the representations for the first ten subjects under the mixture noise scenario with the occlusion percentage is 50%. There are 7×10=70 training images and 12×10=120 testing images. Our representation can preserve well the structure information through representation similarity. Although the training images are severely occluded, we are able to learn robust and discriminative representation. Fig. 16 shows some results of image decomposition on the Extended Yale B database. We also list the results of LSLR [67] method for comparison. From Fig. 16, we can observe that RDLRR can better characterize the occlusion and capture more detailed information of images than LSLR method.

4.4. Parameter discussion

In this part, we will study the impact of λ and α on the recognition performance of the proposed method in different face recognition scenarios. The experimental setting is the same as the above experiments. In our experiments, we just change one parameter while fixing the other one. For face recognition with illumination changes, Subset 4 of the Extended Yale B database is used for training and Subset 5 for testing. For Multi-PIE database, Session 1 with slight illumination changes is used for training and Session 2 for testing. For face recognition with real disguise, 6 images (3 neutral images plus 3 images with sunglasses or scarf) in session 1 are used for training, 3 images with sunglasses or scarf in session 2 are used for testing. For face recognition with occlusions, Subset 1 of the Extended Yale B database is used for training and Subset 2 for testing (occlusion level is 40%).

Fig. 17 shows the recognition rates of RDLRR versus different parameters in different face recognition scenarios. From Fig. 17(a), we can see that RDLRR always achieves better performance when parameter λ is larger than 3. In Fig. 17(b), for face recognition with illumination changes, Sunglasses and Scarf, RDLRR always achieves better performance when parameter α is lower than 0.5. For face recognition with occlusions, RDLRR always achieves stable performance when α varies from 0.1 to 1.

4.5. Limitations

Similar to other representation-based methods for face recognition, registered face images are needed for training and testing in our method. In other words, such approaches cannot directly be applied for recognizing face images with pose variations. Thus, these methods are particularly favorable for applications such as access control, automatic teller machine, or other security facilities. In such scenarios, one can collect controlled (registered) training images in advance, and the test image can be captured under the same (or very similar) environments. However, if the registered face images are not available for either training or testing (but only with shift and in-plane rotation variations), one can apply existing image registration techniques like RASL [78], which can alleviate the above limitations for representation-based approaches.

5. Conclusions

In this paper, we proposed to learn robust and discriminative low-rank representation (RDLRR) for robust face recognition in case that both training and testing images are corrupted due to occlusion. By introducing low-rank assumption to characterize the representation and each error term simultaneously, RDLRR could capture the global structure of data and the holistic structure of each error image. By adding an ideal-code regularization term and the classification error constraint, the learned representation was optimal for classification purpose. After obtaining the robust and discriminative representation, a simple yet powerful linear multi-classifier was used for final classification tasks. Extensive experiments demonstrated that the proposed RDLRR was robust to corruptions: illumination changes, real disguise and block occlusion, and yielded better performances compared to state-of-the-art methods.

Acknowledgement

The authors would like to thank the editor and the anonymous reviewers for their critical and constructive comments and suggestions. This work was partially supported by the National Science Fund for Distinguished Young Scholars under Grand nos. 61125305, 61472187, 61233011 and 61373063, the National Natural Science Foundation of China under Grant nos. 61272273, 61375001, 61401228, 61473086, 61502245, 61503195, 61502081, 61533010, 61502244 and 61503188, the China Postdoctoral Science Foundation under Grant nos. 2015M581841 and 2016M600433, the Natural Science Foundation of Jiangsu Province under Grant nos. BK20150849 and BK20150982, the Key Project of Chinese Ministry of Education under Grant no. 313030, the 973 Program no. 2014CB349303, and Program for Changjiang Scholars and Innovative Research Team in University (No. IRT13072).

References

[1] X. Wen, L. Shao, Y. Xue, W. Fang, A rapid learning algorithm for vehicle classification,, Inf. Sci. 295 (2015) 395–406.
[2] G. Gao, J. Yang, J. Qian, L. Zhang, Integration of multiple orientation and texture information for finger-knuckle-print verification, Neurocomputing 135 (2014) 180–191.
[3] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, K. Ren, A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing, IEEE Trans. Inf. Forensics Secur. 11 (11) (2016) 2594–2608.
[4] J. Li, X. Li, B. Yang, X. Sun, Segmentation-based image copy-move forgery detection scheme, IEEE Trans. Inf. Forensics Secur. 10 (3) (2015) 507–518.
[5] F. Shen, C. Shen, X. Zhou, Y. Yang, H.T. Shen, Face image classification by pooling raw features, Pattern Recognit. 54 (2016) 94–103.
[6] Z. Zhou, Y. Wang, Q. Wu, C. Yang, X. Sun, Effective and efficient global context verification for image copy detection, IEEE Trans. Inf. Forensics Secur. (2016). http://dx.doi.org/10.1109/TIFS.2016.2601065.

[7] S. Barra, A. Casanova, F. Narducci, S. Ricciardi, Ubiquitous iris recognition by means of mobile devices, Pattern Recognit. Lett. 57 (2015) 66–73.
[8] C. Ding, D. Tao, Robust face recognition via multimodal deep face representation, IEEE Trans. Multimed. 17 (11) (2015) 2049–2058.
[9] F. Shen, C. Shen, Q. Shi, A. van den Hengel, Z. Tang, H.T. Shen, Hashing on nonlinear manifolds, IEEE Trans. Image Process. 24 (6) (2015) 1839–1851.
[10] Z. Fu, X. Wu, C. Guan, X. Sun, K. Ren, Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement, IEEE Trans. Inf. Forensics Secur. 11 (12) (2016) 2706–2716.
[11] Z. Xia, X. Wang, X. Sun, Q. Wang, A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data, IEEE Trans. Parallel Distrib. Syst. 27 (2) (2016) 340–352.
[12] Z. Pan, Y. Zhang, S. Kwong, Efficient motion and disparity estimation optimization for low complexity multiview video coding, IEEE Trans. Broadcast. 61 (2) (2015) 166–176.
[13] C. Yuan, X. Sun, R. Lv, Fingerprint liveness detection based on multi-scale LPQ and PCA, China Commun. 13 (7) (2016) 60–65.
[14] Z. Xia, X. Wang, X. Sun, Q. Liu, N. Xiong, Steganalysis of LSB matching using differences between nonadjacent pixels, Multimed. Tools Appl. 75 (4) (2016) 1947–1962.
[15] Y. Zheng, B. Jeon, D. Xu, Q. Wu, H. Zhang, Image segmentation by generalized hierarchical fuzzy C-means algorithm, J. Intell. Fuzzy Syst. 28 (2) (2015) 961–973.
[16] G. Gao, L. Zhang, J. Yang, L. Zhang, D. Zhang, Reconstruction based finger-knuckle-print verification with score level adaptive binary fusion, IEEE Trans. Image Process. 22 (12) (2013) 5050–5062.
[17] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Trans. Circuits Syst. Video Technol. 14 (1) (2004) 4–20.
[18] Z. Lai, Y. Xu, Q. Chen, J. Yang, D. Zhang, Multilinear sparse principal component analysis, IEEE Trans. Neural Netw. Learn. Syst. 25 (10) (2014) 1942–1950.
[19] Z. Lai, W.K. Wong, Y. Xu, C. Zhao, M. Sun, Sparse alignment for robust tensor learning, IEEE Trans. Neural Netw. Learn. Syst. 25 (10) (2014) 1779–1792.
[20] W.K. Wong, Z.-H. Lai, Y. Xu, J. Wen, C.P. Ho, Joint tensor feature analysis for visual object recognition, IEEE Trans. Cybern. 45 (11) (2015) 2425–2436.
[21] M. Yang, P. Zhu, F. Liu, L. Shen, Joint representation and pattern learning for robust face recognition, Neurocomputing 168 (2015) 70–80.
[22] Z. Lai, W.K. Wong, Y. Xu, J. Yang, D. Zhang, Approximate orthogonal sparse embedding for dimensionality reduction, IEEE Trans. Neural Netw. Learn. Syst. 27 (4) (2016) 723–735.
[23] B. Chen, H. Shu, G. Coatrieux, G. Chen, X. Sun, J.L. Coatrieux, Color image analysis by quaternion-type moments,, J. Math. Imaging Vis. 51 (1) (2015) 124–144.
[24] W. Yang, Z. Wang, C. Sun, A collaborative representation based projections method for feature extraction, Pattern Recognit. 48 (1) (2015) 20–27.
[25] R. He, Y. Zhang, Z. Sun, Q. Yin, Robust subspace clustering with complex noise, IEEE Trans. Image Process. 24 (11) (2015) 4001–4013 (Nov).
[26] R. He, W.-S. Zheng, B.-G. Hu, X.-W. Kong, Two-stage nonnegative sparse representation for large-scale face recognition, IEEE Trans. Neural Netw. Learn. Syst. 24 (1) (2013) 35–46.
[27] J. Gui, Z. Sun, W. Jia, R. Hu, Y. Lei, S. Ji, Discriminant sparse neighborhood preserving embedding for face recognition, Pattern Recognit. 45 (8) (2012) 2884–2893.
[28] P. Huang, C. Chen, Z. Tang, Z. Yang, Feature extraction using local structure preserving discriminant analysis, Neurocomputing 140 (2014) 104–113.
[29] H. Nguyen, W. Yang, F. Shen, C. Sun, Kernel Low-Rank Representation for face recognition, Neurocomputing 155 (2015) 32–42.
[30] P. Huang, G. Gao, Parameterless reconstructive discriminant analysis for feature extraction, Neurocomputing 190 (2016) 50–59.
[31] X.-Y. Jing, F. Wu, X. Zhu, X. Dong, F. Ma, Z. Li, Multi-spectral low-rank structured dictionary learning for face recognition, Pattern Recognit. 59 (2016) 14–25.
[32] C. Xiong, L. Liu, X. Zhao, S. Yan, T.-K. Kim, Convolutional fusion network for face verification in the wild, IEEE Trans. Circuits Syst. Video Technol. 26 (3) (2016) 517–528.
[33] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via Sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[34] I. Naseem, R. Togneri, M. Bennamoun, Linear Regression for Face Recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 2106–2112.
[35] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, Y. Ma, Toward a practical face recognition system: Robust alignment and illumination by sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2) (2012) 372–386.
[36] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: Which helps face recognition? in: Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), 2011, pp. 471–478.
[37] B. Gu, V.S. Sheng, K.Y. Tay, W. Romano, S. Li, Incremental support vector learning for ordinal regression, IEEE Trans. Neural Netw. Learn. Syst. 26 (7) (2015) 1403–1416.
[38] B. Gu, V.S. Sheng, Z. Wang, D. Ho, S. Osman, S. Li, Incremental learning for v-support vector regression, Neural Netw. 67 (2015) 140–150.
[39] G. Gao, J. Yang, S. Wu, X. Jing, D. Yue, Bayesian sample steered discriminative regression for biometric image classification, Appl. Soft Comput. 37 (2015) 48–59.
[40] B. Gu, V.S. Sheng, A Robust regularization path algorithm for v-support vector classification, IEEE Trans. Neural Netw. Learn. Syst. (2016). http://dx.doi.org/10.1109/TNNLS.2016.2527796.
[41] B. Gu, X. Sun, V.S. Sheng, Structural Minimax probability machine, IEEE Trans. Neural Netw. Learn. Syst. (2016). http://dx.doi.org/10.1109/TNNLS.2016.2544779.
[42] F. Yin, L. Jiao, F. Shang, L. Xiong, S. Mao, Double linear regressions for single labeled image per person face recognition, Pattern Recognit. 47 (4) (2014) 1547–1558.
[43] L. Luo, J. Yang, J. Qian, Y. Tai, Nuclear-L1 norm joint regression for face reconstruction and recognition with mixed noise, Pattern Recognit. 48 (12) (2015) 3811–3824.
[44] F. Shen, C. Shen, A. van den Hengel, Z. Tang, Approximate least trimmed sum of squares fitting and applications in image analysis, IEEE Trans. Image Process. 22 (5) (2013) 1836–1847.
[45] G. Gao, J. Yang, A novel sparse representation based framework for face image super-resolution, Neurocomputing 134 (2014) 92–99.
[46] A. Tawari, M.M. Trivedi, Face expression recognition by cross modal data association, IEEE Trans. Multimed. 15 (7) (2013) 1543–1552.
[47] B. Ma, J. Shen, Y. Liu, H. Hu, L. Shao, X. Li, Visual tracking using strong classifier and structural local sparse descriptors, IEEE Trans. Multimed. 17 (10) (2015) 1818–1828.
[48] J. Yang, L. Zhang, Y. Xu, J.-y. Yang, Beyond sparsity: the role of L 1-optimizer in pattern classification, Pattern Recognit. 45 (3) (2012) 1104–1118.
[49] M. Yang, L. Zhang, J. Yang, D. Zhang, Robust Sparse Coding for Face Recognition, in: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun 2011, pp. 625–632.
[50] R. He, W.-S. Zheng, T. Tan, Z. Sun, Half-quadratic-based iterative minimization for robust sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 261–275.
[51] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, Y. Xu, Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes, IEEE Trans. Pattern Anal. Mach. Intell. 39 (1) (2017) 156–171.
[52] L. Luo, J. Yang, J. Qian, J. Yang, Nuclear Norm Regularized Sparse Coding, in: Proceedings of the 2014 22nd International Conference on Pattern Recognition (ICPR), 2014, pp. 1834–1839.
[53] F. Zhang, J. Yang, Y. Tai, J. Tang, Double nuclear norm-based matrix decomposition for occluded image recovery and background modeling, IEEE Trans. Image Process. 24 (6) (2015) 1956–1966.
[54] J. Qian, L. Luo, J. Yang, F. Zhang, Z. Lin, Robust nuclear norm regularized regression for face recognition with occlusion, Pattern Recognit. 48 (10) (2015) 3145–3159.
[55] X.H. Shen, Y. Wu, A Unified Approach to Salient Object Detection via Low Rank Matrix Recovery, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun 2012, pp. 853–860.
[56] X.Y. Cui, J.Z. Huang, S.T. Zhang, D.N. Metaxas, Background Subtraction Using Low Rank and Group Sparsity Constraints, in: Proceedings of the 2012 European Conference on Computer Vision (ECCV), Oct 2012, pp. 612–625.
[57] T.Z. Zhang, B. Ghanem, S. Liu, N.Ahuja, Low-Rank Sparse Learning for Robust Visual Tracking, in: Proceedings of the 2012 European Conference on Computer Vision (ECCV), Oct 2012, pp. 470–484.
[58] S. Gu, L. Zhang, W. Zuo, X. Feng, Weighted nuclear norm minimization with application to image denoising, in: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2862–2869.
[59] C.F. Chen, C.P. Wei, Y.C.F. Wang, Low-Rank Matrix Recovery with Structural Incoherence for Robust Face Recognition, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun, 2012, pp. 2618–2625.
[60] C.-P. Wei, C.-F. Chen, Y.-C.F. Wang, Robust face recognition With Structurally incoherent low-Rank matrix decomposition, IEEE Trans. Image Process. 23 (8) (2014) 3294–3307.
[61] E.J. Candes, X.D. Li, Y. Ma, J. Wright, Robust principal component analysis? J. ACM 58 (3) (2011) 1–37.
[62] J. Wright, A. Ganesh, S. Rao, Y. Peng, Y. Ma, Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization, Adv. Neural Inf. Process. Syst. (2009) 2080–2088.
[63] L. Ma, C.H. Wang, B.H. Xiao, W. Zhou, Sparse Representation for Face Recognition based on Discriminative Low-Rank Dictionary Learning, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun 2012, pp. 2586–2593.
[64] G.C. Liu, Z.C. Lin, S.C. Yan, J. Sun, Y. Yu, Y. Ma, Robust Recovery of Subspace Structures by Low-Rank Representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 171–184.
[65] G.C. Liu, S.C. Yan, Latent Low-Rank Representation for Subspace Segmentation and Feature Extraction, in: Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Nov 2011, pp. 1615–1622.
[66] Y. M, C. S. T, G. J. B, Robust face recognition via double low-rank matrix recovery for feature extraction,2013 in: Proceedings of the 20th IEEE International Conference on Image Processing (ICIP), Sep 2013, pp. 3770–3774.
[67] Y. Zhang, Z. Jiang, L.S. Davis, Learning structured low-rank representations for image classification, in: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 676–683.
[68] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, pp. 586–591.
[69] N. Kwak, Principal component analysis based on L1-norm maximization, IEEE Trans. Pattern Anal. Mach. Intell. 30 (9) (2008) 1672–1680.
[70] Z. Jiang, Z. Lin, L.S. Davis, Label consistent k-svd: learning a discriminative dictionary for recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (11) (2013) 2651–2664.
[71] K.C. Lee, J. Ho, D.J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 684–698.

[72] Z. Lin, M. Chen, and Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, arXiv preprint arXiv:1009.5055, 2010.
[73] J.F. Cai, E.J. Candes, Z.W. Shen, A singular value thresholding algorithm for matrix Completion, Siam J. Optim. 20 (4) (2010) 1956–1982.
[74] Y. Zhang, Recent advances in alternating direction methods: Practice and theory, in: Proceedings of the IPAM Workshop: Numerical Methods for Continuous Optimization. UCLA, Los Angeles, 2010.
[75] J. Eckstein, D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal Monotone-operators, Math. Program. 55 (3) (1992) 293–318.
[76] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, Image Vis. Comput. 28 (5) (2010) 807–813.
[77] A.M. Martinez, R. Benavente, The AR face database, CVC Tech. Rep. 24 (1998).
[78] Y. Peng, A. Ganesh, J. Wright, W. Xu, Y. Ma, RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2233–2246.

Guangwei Gao received the B.S. degree in information and computation science from Nanjing Normal University, Nanjing, China, in 2009, and the Ph.D. degree in pattern recognition and intelligence systems from Nanjing University of Science and Technology, Nanjing, China, in 2014. From March 2011 to September 2011 and February 2013 to August 2013, he was an exchange student of Department of Computing, Hong Kong Polytechnic University. Now, he is an assistant professor in the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications. His research interests include face recognition, face hallucination and biometrics.

Jian Yang received the BS degree in mathematics from the Xuzhou Normal University in 1995. He received the MS degree in applied mathematics from the Changsha Railway University in 1998 and the Ph.D. degree from the Nanjing University of Science and Technology (NUST), on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a postdoctoral researcher at the University of Zaragoza. From 2004 to 2006, he was a Postdoctoral Fellow at Biometrics Centre of Hong Kong Polytechnic University. From 2006 to 2007, he was a Postdoctoral Fellow at Department of Computer Science of New Jersey Institute of Technology. Now, he is a professor in the School of Computer Science and Technology of NUST. He is the author of more than 80 scientific papers in pattern recognition and computer vision. His journal papers have been cited more than 3000 times in the ISI Web of Science, and 7000 times in the Web of Scholar Google. His research interests include pattern recognition, computer vision and machine learning. Currently, he is an associate editor of Pattern Recognition Letters and IEEE Trans. Neural Networks and Learning Systems, respectively.

Xiaoyuan Jing received the M.Sc. and Ph.D. degrees in pattern recognition and intelligence systems from the Nanjing University of Science and Technology, Nanjing, China, in 1995 and 1998, respectively. He was an Associate Researcher with the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2002. He was a Professor with the Shenzhen Graduate School, Harbin Institute of Technology, Harbin, China, in 2005. He has been a Research Fellow with Hong Kong Polytechnic University, Kowloon, Hong Kong. He is currently a Professor and Doctoral Supervisor with the State Key Laboratory of Software Engineering, Wuhan University, Wuhan, China, and with the College of Automation, Nanjing University of Posts and Telecommunications, Nanjing. He has authored and coauthored more than 30 scientific papers in SCI indexed international journals, such as the IEEE Transactions on Pattern Recognition, Signal Processing, Pattern Recognition Letters, and Neurocomputing, more than 20 papers in international conferences, and a book on biometrics from American IGP Press. His current research interests include pattern recognition, machine learning, data mining, image processing, neural networks, and artificial intelligence. Currently he is a reviewer of more than 20 international journals and conferences. He was awarded the New Century Excellent Talent Award by the Chinese Education Ministry.

Fumin Shen received his Bachelor degree at 2007 and Ph.D. degree at 2014 from Shandong University and Nanjing University of Science and Technology, China, respectively. Now he is a lecturer of University of Electronic Science and Technology of China. His major research interests include computer vision and machine learning, including face recognition, image analysis and hashing methods.

Wankou Yang received his B.S., M.S., and Ph.D. degrees from the School of Computer Science and Technology, Nanjing University of Science and Technology, China, in 2002, 2004, and 2009, respectively. He is currently an Assistant Professor in School of Automation, Southeast University, Nanjing, China. He has published over 60 high-quality research articles on the professional journals and conferences. His research interests include pattern recognition, image processing and machine learning.

Dong Yue received the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 1995. He is currently a Professor and the Dean with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China, and also a Changjiang Professor with the Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan, China. His current research interests include analysis and synthesis of networked control systems, multiagent systems, optimal control of power systems, and internet of things. He has published over 100 papers in international journals, domestic journals, and international conferences. Prof. Yue is currently an Associate Editor of the IEEE Control Systems Society Conference Editorial Board and the International Journal of Systems Science.