Pattern Analysis and Applications
https://doi.org/10.1007/s10044-019-00858-4

SHORT PAPER

Learning a representation with the block-diagonal structure for pattern classification

He-Feng Yin[1] · Xiao-Jun Wu[1] · Josef Kittler[2] · Zhen-Hua Feng[2]

Received: 7 December 2018 / Accepted: 22 November 2019

Abstract
Sparse-representation-based classification (SRC) has been widely studied and developed for various practical signal classification applications. However, the performance of a SRC-based method is degraded when both the training and test data are corrupted. To counteract this problem, we propose an approach that learns representation with block-diagonal structure (RBDS) for robust image recognition. To be more specific, we first introduce a regularization term that captures the block-diagonal structure of the target representation matrix of the training data. The resulting problem is then solved by an optimizer. Last, based on the learned representation, a simple yet effective linear classifier is used for the classification task. The experimental results obtained on several benchmarking datasets demonstrate the efficacy of the proposed RBDS method. The source code of our proposed RBDS is accessible at https://github.com/yinhefeng/RBDS.

Keywords Pattern classification · Low-rank and sparse representation · Block-diagonal structure

1 Introduction

In recent years, sparse representation has gained significant attention due to its successful applications in image recognition [1, 2], object tracking [3], subspace clustering [4] and many other computer vision tasks [5–7]. The pioneering work in pattern recognition utilizing sparse-representation-based classification (SRC) is attributed to Wright et al. [8]. SRC expresses an input test pattern as a sparse linear superposition of all the training data. Its classification is performed by checking which class conditional subset of the reconstruction coefficients produces the lowest reconstruction error.

✉ Xiao-Jun Wu
wu_xiaojun@jiangnan.edu.cn

He-Feng Yin
yinhefeng@126.com

Josef Kittler
j.kittler@surrey.ac.uk

Zhen-Hua Feng
z.feng@surrey.ac.uk

[1] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China

[2] Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford GU2 7XH, UK

It has been demonstrated in [8] that SRC is robust when the test image is occluded or corrupted, provided the training data are clean (i.e., no occlusion or corruption). However, when the training data contain occluded or corrupted samples, the performance of SRC is degraded. In this paper, we address the problem when both the training and test data are corrupted and present an approach that alleviates it.

To improve the robustness of SRC with respect to corrupted training data, a low-rank matrix recovery (LRMR) method has been proposed to obtain a low-rank part of the corrupted image content. There are a number of previous attempts to deal with outliers. For instance, Candes et al. [9] presented the robust PCA (RPCA), which assumes that the observations lie in a single subspace such that they can be decomposed into two separate components, i.e., the low-rank normal data and a sparse noise part. However, RPCA cannot handle the situation where corrupted or outlying data are drawn from a union of multiple subspaces. To this end, Liu et al. [10] proposed a low-rank representation (LRR) method.

Based on LRMR, many approaches have been presented for robust pattern classification. Ma et al. [11] presented a discriminative low-rank dictionary for sparse representation (DLRD_SR). By integrating rank minimization into sparse representation for dictionary learning, this method achieved impressive face recognition results, especially in

Published online: 12 December 2019

the presence of corruption. Zhang et al. [12] proposed a low-rank structure representation for image classification by adding an ideal-code regularization term to the objective function.

Recently, Li et al. [13] advocated discriminative dictionary learning with low-rank regularization ($D^2L^2R^2$) for image classification that can handle training samples corrupted with large noise. $D^2L^2R^2$ combines the Fisher discrimination function with a low-rank constraint on the sub-dictionary to make the learned dictionary more discerning and pure. Inspired by the low-rank constraint on the sub-dictionary and the ideal-code regularization term, Nguyen et al. [14] proposed a discriminative low-rank dictionary learning (DLR_DL) method for face recognition.

Zheng et al. [15] designed a novel low-rank matrix recovery algorithm with the Fisher discriminant regularization (FDLR). Wei et al. [16] introduced a constraint of structural incoherence into RPCA and presented a method called low-rank matrix recovery with structural incoherence (LRSI). Based on LRSI, Yin et al. [17] presented a new method that can correct the corrupted test images with a low-rank projection matrix. Later, Chen et al. [18] proposed a discriminative low-rank representation (DLRR) method by incorporating structural incoherence into the framework of LRR.

Dong et al. [19] explored a discriminative orthonormal dictionary learning method for low-rank representation. Rong et al. [20] presented a novel low-rank double dictionary learning (LRD^2L) approach that simultaneously learns a low-rank class-specific sub-dictionary for each class and a low-rank class-shared dictionary. Gao et al. [21] constructed a robust and discriminative low-rank representation (RDLRR) by exploiting the low-rank characteristics of both the data representation and each occlusion-induced error image simultaneously. Du et al. [22] introduced a discriminative low-rank graph preserving dictionary learning (DLRGP_DL) method to learn a discriminative structured dictionary for sparse-representation-based image recognition. Recently, Wu et al. [23] proposed a gradient direction-based hierarchical adaptive sparse and low-rank (GD-HASLR) model to solve the real-world occluded face recognition problem.

Though the aforementioned methods achieve encouraging results in various classification tasks, structural information content of the training data is not fully exploited. Indeed, all recent works [24, 25] indicate that utilizing structural information can achieve better performance in recognition tasks. In [25], there is no dictionary learning process, i.e., the training data are directly utilized as the dictionary. This impacts on the classification performance that will be adversely affected when both the training and test data are corrupted.

In this paper, aiming at overcoming the above drawbacks, we propose an approach that learns a representation with block-diagonal structure (RBDS) for robust recognition. Concretely, a regularization term, which can capture the block-diagonal structure of the target representation matrix of the training data, is introduced, enhancing the discriminative potential of the learned representations. Furthermore, we adopt an innovative strategy to solve the resulting optimization problem. In addition, a compact dictionary is learned by our approach.

In summary, our main contributions include:

1. A new approach that learns a robust representation mirroring a block-diagonal structure is developed, which is insensitive to corruption of both training and testing images.
2. A compact dictionary with favorable reconstruction and discrimination properties is learned in the training stage of our proposed method.
3. An effective optimization technique based on the alternating direction method of multipliers (ADMM) is presented to solve the proposed problem.

The remainder of this paper is structured as follows. Section 2 reviews related work on low-rank matrix recovery. In Sect. 3, we present our proposed approach, with detailed optimization procedures given in Sect 4. Section 5 reports the experimental results on five benchmarking datasets. Last, the conclusion is drawn in Sect. 6.

2 Low-rank matrix recovery

Suppose the data matrix $\mathbf{X}$ can be decomposed into two matrices, i.e., $\mathbf{X} = \mathbf{A} + \mathbf{E}$, where $\mathbf{A}$ is a low-rank matrix and $\mathbf{E}$ is the error matrix. The robust principal component analysis (RPCA) derives a low-rank matrix $\mathbf{A}$ from the corrupted data matrix $\mathbf{X}$ [9]. The objective function of RPCA is formulated as,

$$\min_{\mathbf{A},\mathbf{E}} rank(\mathbf{A}) + \lambda \|\mathbf{E}\|_0, \text{ s.t. } \mathbf{X} = \mathbf{A} + \mathbf{E} \tag{1}$$

where $rank(\cdot)$ is the rank of a matrix, $\|\cdot\|_0$ means the ℓ_0 pseudo-norm and λ is a balance parameter. RPCA implicitly assumes that the underlying data structure lies in a single low-rank subspace. However, this assumption is not realistic in many practical applications. Let us take face images as an example. While images of one individual tend to be drawn from the same subspace, images of distinct persons are drawn from different subspaces. Therefore, a more realistic assumption is that data samples are drawn from a union of multiple subspaces.

In order to accommodate data from multiple subspaces, Liu et al. [10] generalized the concept of RPCA and proposed a more general rank minimization problem, which is formulated as,

$$\min_{\mathbf{Z},\mathbf{E}} rank(\mathbf{Z}) + \lambda\|\mathbf{E}\|_0 \text{ s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \tag{2}$$

where $\mathbf{D}$ is a dictionary that spans the data space.

3 The proposed method

We first introduce the notations to be used in this paper. $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_C] \in \mathbb{R}^{d\times n}$ is the matrix of training data from C classes, $\mathbf{X}_i \in \mathbb{R}^{d\times n_i}$ is the matrix of class i with n_i samples of dimension d and $n = \sum_{i=1}^{C} n_i$. $\mathbf{1}_m = [1, \ldots, 1]^T \in \mathbb{R}^{m\times 1}$ denotes the all-one vector. Each sample in $\mathbf{X}$ can be expressed by the linear superposition of atoms in dictionary $\mathbf{D}$,

$$\mathbf{X} = \mathbf{DZ} \tag{3}$$

where $\mathbf{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_m] \in \mathbb{R}^{d\times m}$ is the dictionary, $\boldsymbol{d}_i$ is the i-th atom in $\mathbf{D}$ and $\mathbf{Z} \in \mathbb{R}^{m\times n}$ is the representation matrix of the training data.

3.1 Low-rank and sparse representation

It has been convincingly demonstrated in the literature that sparse representation achieves promising results in classification tasks. Similarly, it has been established that the low-rank property is a powerful concept enabling to capture the structure information of high-dimensional data, which is robust to sparse noise. In [12], low-rank and sparse representation are combined to exploit the above two aspects. Accordingly, the problem of learning low-rank and sparse representation can be formulated as follows,

$$\min_{\mathbf{Z},\mathbf{E}} rank(\mathbf{Z}) + \lambda\|\mathbf{E}\|_0 + \beta\|\mathbf{Z}\|_0 \text{ s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \tag{4}$$

Due to the discrete properties of the rank function and the ℓ_0-norm minimization, it is practically difficult to solve Eq. (4). A common way is to replace the rank function and ℓ_0-norm with nuclear norm and ℓ_1-norm, respectively. Thus, Eq. (4) can be reformulated as,

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 + \beta\|\mathbf{Z}\|_1 \text{ s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \tag{5}$$

where λ and β control the sparsity of the noise term $\mathbf{E}$ and of the representation term $\mathbf{Z}$. The low-rank and sparse representation can be obtained by solving Eq. (5) with respect to the given dictionary $\mathbf{D}$.

The results reported in [12] demonstrate the effectiveness of the low-rank and sparse representation for classification tasks. However, the solution does not focus on discriminative information. To rectify this deficiency, in the next section, we present our approach which learns a representation mirroring the block-diagonal structure (RBDS) for pattern classification.

3.2 Exploiting the block-diagonal structure

Several works have exploited the block-diagonal structure of the representation matrix $\mathbf{Z}$ [12, 24, 25]. For example, Zhang et al. [12] proposed a discriminative, low-rank structure framework for image classification by introducing an idealized structure as a regularization constraint. Accordingly, the influence of the samples from the same class on an input pattern representation is regularized to be the same. Later, Li et al. [24] argued that it is unreasonable to introduce such an ideal regularization term. They presented an algorithm to learn Representation with a classwise block-diagonal (RCBD) structure. Recently, Zhang et al. [25] developed a discriminative block-diagonal low-rank representation (BDLRR) for recognition.

Our proposed approach is similar to BDLRR. However, BDLRR does not involve dictionary learning, and this oversight inhibits the method to realize its potential when both the training and test data are corrupted. We propose an intuitive way to exploit the block-diagonal structure inherent in the training data to minimize the off-block-diagonal entries of the representation matrix. We seek to capture the block-diagonal structure by adding a regularization term $\|\mathbf{A} \odot \mathbf{Z}\|_F^2$, where $\mathbf{A}$ is defined as,

$$\mathbf{A}(i,j) = \begin{cases} 0, & \text{if } \boldsymbol{d}_i \text{ and } \boldsymbol{x}_i \text{ belong to the same class} \\ 1, & \text{otherwise} \end{cases} \tag{6}$$

in which $\boldsymbol{d}_i$ is the i-th atom in dictionary $\mathbf{D}$. An example of $\mathbf{A}$ is shown in Fig. 1. Now, the representation learning problem with block-diagonal regularization term can be formulated as follows,

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 + \frac{\alpha}{2}\|\mathbf{A} \odot \mathbf{Z}\|_F^2 + \beta\|\mathbf{Z}\|_1, \text{ s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \tag{7}$$

where λ, α and β are the balancing parameters for each component and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

3.3 Dictionary learning

Compact and discriminative dictionary plays an important role in robust pattern classification, especially when both the training and test data are corrupted due to occlusion and pixel corruption. As dictionary learning has been proven to achieve promising performance [11, 16], we incorporate it into our proposed framework. Accordingly, the final formulation of our proposed approach can be stated as follows,

$$\min_{\mathbf{Z},\mathbf{E},\mathbf{D}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 + \frac{\alpha}{2}\|\mathbf{A} \odot \mathbf{Z}\|_F^2 + \beta\|\mathbf{Z}\|_1 + \frac{\gamma}{2}\|\mathbf{D}\|_F^2, \text{ s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \tag{8}$$

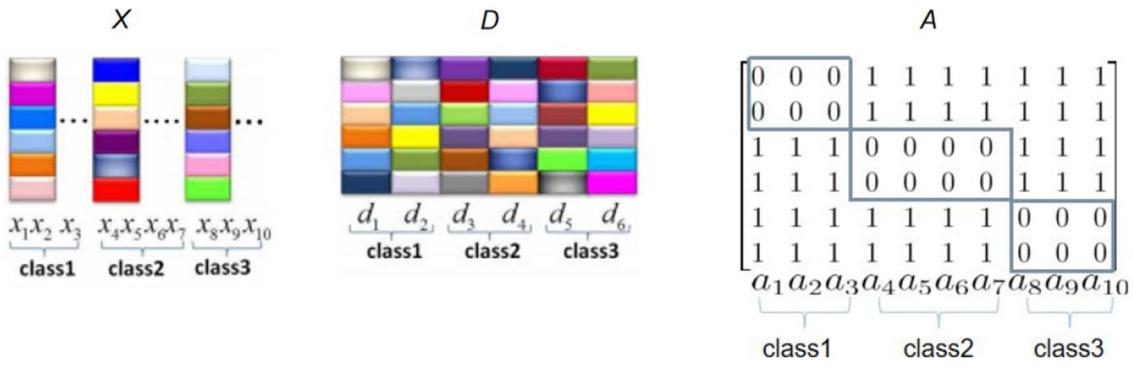


Fig. 1 An illustration of **A**

where $\gamma\|\mathbf{D}\|_F^2$ is to prevent a scale change during the dictionary learning process.

3.4 Classification based on RBDS

The outcome of the training phase of RBDS is a dictionary, $\mathbf{D}$, and the representation matrix, $\mathbf{Z}$, of the training data $\mathbf{X}$. For the test data $\mathbf{X}_{test}$, we obtain the corresponding representation matrix $\hat{\mathbf{Z}}$ by solving Eq. (5), where $\hat{z}_j$ is the representation vector of the j-th test sample. We employ a simple linear classifier to perform recognition. The linear classifier $\mathbf{W}^*$ is designed using the representation $\mathbf{Z}$ of the training data and its label matrix $\mathbf{H}$. The problem of learning $\mathbf{W}^*$ can be formulated as follows,

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \|\mathbf{H} - \mathbf{W}\mathbf{Z}\|_F^2 + \eta\|\mathbf{W}\|_F^2 \tag{9}$$

where $\eta > 0$ is a parameter. It is easy to obtain the following closed-form solution for Eq. (9),

$$\mathbf{W}^* = \mathbf{H}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T + \eta\mathbf{I})^{-1} \tag{10}$$

Then the identity of a test sample j is determined by,

$$i^* = \arg\max_i \mathbf{W}^*\hat{z}_j \tag{11}$$

where i^* corresponds to the largest output.

4 Optimization algorithm

To solve the optimization problem Eq. (8), we obtain the following equivalent problem by introducing two auxiliary variables $\mathbf{J}$ and $\mathbf{L}$. Then Eq. (8) can be rewritten as,

$$\min_{\mathbf{Z},\mathbf{J},\mathbf{L},\mathbf{E},\mathbf{D}} \|\mathbf{J}\|_* + \lambda\|\mathbf{E}\|_1 + \frac{\alpha}{2}\|\mathbf{A}\odot\mathbf{Z}\|_F^2 + \beta\|\mathbf{L}\|_1 + \frac{\gamma}{2}\|\mathbf{D}\|_F^2,$$
$$\text{s.t. } \mathbf{X} = \mathbf{D}\mathbf{Z} + \mathbf{E}, \mathbf{Z} = \mathbf{J}, \mathbf{Z} = \mathbf{L} \tag{12}$$

which can be solved based on the augmented Lagrange multiplier (ALM) method [26]. The augmented Lagrangian function of Eq. (12) is defined as follows,

$$\begin{aligned}
&\Lambda(\mathbf{Z},\mathbf{J},\mathbf{L},\mathbf{E},\mathbf{D},\mathbf{Y}_1,\mathbf{Y}_2,\mathbf{Y}_3,\mu)\\
&= \|\mathbf{J}\|_* + \lambda\|\mathbf{E}\|_1 + \frac{\alpha}{2}\|\mathbf{A}\odot\mathbf{Z}\|_F^2 + \beta\|\mathbf{L}\|_1\\
&\quad + \frac{\gamma}{2}\|\mathbf{D}\|_F^2\\
&\quad + <\mathbf{Y}_1, \mathbf{X} - \mathbf{D}\mathbf{Z} - \mathbf{E}> + <\mathbf{Y}_2, \mathbf{Z} - \mathbf{J}> + <\mathbf{Y}_3, \mathbf{Z} - \mathbf{L}>\\
&\quad + \frac{\mu}{2}(\|\mathbf{X} - \mathbf{D}\mathbf{Z} - \mathbf{E}\|_F^2 + \|\mathbf{Z} - \mathbf{J}\|_F^2 + \|\mathbf{Z} - \mathbf{L}\|_F^2)
\end{aligned} \tag{13}$$

where $<\mathbf{A},\mathbf{B}>= trace(\mathbf{A}^t\mathbf{B})$, $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{Y}_3$ are Lagrange multipliers and $\mu > 0$ is a penalty parameter. The optimization of Eq. (13) can be solved iteratively by updating $\mathbf{J}$, $\mathbf{Z}$, $\mathbf{L}$, $\mathbf{E}$ and $\mathbf{D}$ once at a time. The detailed updating procedures are presented as follows.

Updating $\mathbf{J}$: Fix the other variables and update $\mathbf{J}$ by solving the following problem,

$$\begin{aligned}
\mathbf{J}^{k+1} &= \arg\min_{\mathbf{J}} \|\mathbf{J}\|_* + <\mathbf{Y}_2^k, \mathbf{Z}^k - \mathbf{J}> + \frac{\mu^k}{2}\left\|\mathbf{Z}^k - \mathbf{J}\right\|_F^2\\
&= \arg\min_{\mathbf{J}} \frac{1}{\mu^k}\|\mathbf{J}\|_* + \frac{1}{2}\left\|\mathbf{J} - (\mathbf{Z}^k + \frac{\mathbf{Y}_2^k}{\mu^k})\right\|_F^2\\
&= \mathbf{U}S_{\frac{1}{\mu^k}}[\mathbf{\Sigma}]\mathbf{V}^T
\end{aligned} \tag{14}$$

where $(\mathbf{U},\mathbf{\Sigma},\mathbf{V}^T) = SVD(\mathbf{Z}^k + \mathbf{Y}_2^k/\mu^k)$ and $S_\varepsilon[\cdot]$ is the soft-thresholding (shrinkage) operator defined as follows [26],

$$S_\varepsilon[x] = \begin{cases} x-\varepsilon, \text{if } x > \varepsilon \\ x+\varepsilon, \text{if } x < -\varepsilon \\ 0, \text{otherwise} \end{cases} \tag{15}$$

Updating $\mathbf{Z}$: To update $\mathbf{Z}$, we fix all the variables other than $\mathbf{Z}$ and solve the following problem:

$$\begin{aligned} \mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} \frac{\alpha}{2}\|\mathbf{A}\odot\mathbf{Z}\|_F^2 + \frac{\mu^k}{2}\left\|\mathbf{X}-\mathbf{D}^k\mathbf{Z}-\mathbf{E}^k+\frac{\mathbf{Y}_1^k}{\mu^k}\right\|_F^2 \\ + \frac{\mu^k}{2}\left\|\mathbf{Z}-\mathbf{J}^{k+1}+\frac{\mathbf{Y}_2^k}{\mu^k}\right\|_F^2 + \frac{\mu^k}{2}\left\|\mathbf{Z}-\mathbf{L}^k+\frac{\mathbf{Y}_3^k}{\mu^k}\right\|_F^2 \end{aligned} \tag{16}$$

which is equivalent to

$$\begin{aligned} \mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} \frac{\alpha}{2}\|\mathbf{Z}-\mathbf{R}\|_F^2 + \frac{\mu^k}{2}\left\|\mathbf{X}-\mathbf{D}^k\mathbf{Z}-\mathbf{E}^k+\frac{\mathbf{Y}_1^k}{\mu^k}\right\|_F^2 \\ + \frac{\mu^k}{2}\left\|\mathbf{Z}-\mathbf{J}^{k+1}+\frac{\mathbf{Y}_2^k}{\mu^k}\right\|_F^2 + \frac{\mu^k}{2}\left\|\mathbf{Z}-\mathbf{L}^k+\frac{\mathbf{Y}_3^k}{\mu^k}\right\|_F^2 \end{aligned} \tag{17}$$

where $\mathbf{R} = \mathbf{M}\odot\mathbf{Z}^k$ and $\mathbf{M} = \mathbf{1}_m\mathbf{1}_n^T - \mathbf{A}$. Eq. (17) has a closed-form solution, given by

$$\begin{aligned} \mathbf{Z}^{k+1} = \left[(\mathbf{D}^k)^T\mathbf{D}^k + \left(\frac{\alpha}{\mu^k}+2\right)\mathbf{I}\right]^{-1}[(\mathbf{D}^k)^T(\mathbf{X}-\mathbf{E}^k) \\ + \mathbf{J}^{k+1} + \mathbf{L}^k + \frac{\alpha\mathbf{R} + (\mathbf{D}^k)^T\mathbf{Y}_1^k - \mathbf{Y}_2^k - \mathbf{Y}_3^k}{\mu^k}] \end{aligned} \tag{18}$$

Updating $\mathbf{L}$: When we fix the other variables, Eq. (13) degenerates into a function of $\mathbf{L}$, that is,

$$\begin{aligned} \mathbf{L}^{k+1} &= \arg\min_{\mathbf{L}} \frac{\beta}{\mu^k}\|\mathbf{L}\|_1 + \frac{1}{2}\left\|\mathbf{L}-(\mathbf{Z}^{k+1}+\frac{\mathbf{Y}_3^k}{\mu^k})\right\|_F^2 \\ &= S_{\frac{\beta}{\mu^k}}[\mathbf{Z}^{k+1}+\frac{\mathbf{Y}_3^k}{\mu^k}] \end{aligned} \tag{19}$$

Updating $\mathbf{E}$: To update $\mathbf{E}$, we minimize Eq. (13) and fix all the variables other than $\mathbf{E}$, which leads to

$$\begin{aligned} \mathbf{E}^{k+1} &= \arg\min_{\mathbf{E}} \frac{\lambda}{\mu^k}\|\mathbf{E}\|_1 + \frac{1}{2}\left\|\mathbf{E}-(\mathbf{X}-\mathbf{D}^k\mathbf{Z}^{k+1}+\frac{\mathbf{Y}_1^k}{\mu^k})\right\|_F^2 \\ &= S_{\frac{\lambda}{\mu^k}}[\mathbf{X}-\mathbf{D}^k\mathbf{Z}^{k+1}+\frac{\mathbf{Y}_1^k}{\mu^k}] \end{aligned} \tag{20}$$

Updating $\mathbf{D}$: When the other variables are fixed, optimizing Eq. (13) with respect to $\mathbf{D}$ boils down to the following problem,

$$\begin{aligned} \mathbf{D}^{k+1} = \arg\min_{\mathbf{D}} \frac{\gamma}{2}\|\mathbf{D}\|_F^2 + <\mathbf{Y}_1^k, \mathbf{X}-\mathbf{D}\mathbf{Z}^{k+1}-\mathbf{E}^{k+1}> \\ + \frac{\mu^k}{2}\left\|\mathbf{X}-\mathbf{D}\mathbf{Z}^{k+1}-\mathbf{E}^{k+1}\right\|_F^2 \end{aligned} \tag{21}$$

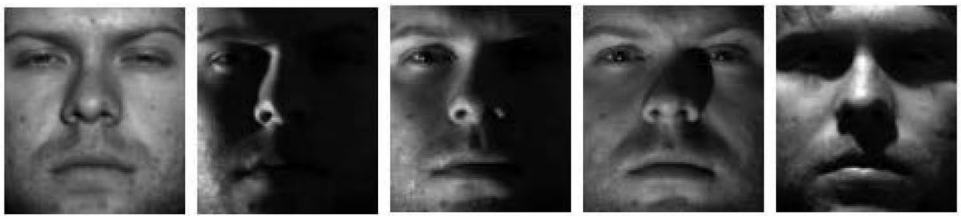
which has a closed-form solution as follows,

$$\begin{aligned} \mathbf{D}^{k+1} = \left[\frac{\mathbf{Y}_1^k(\mathbf{Z}^{k+1})^T}{\mu^k} - (\mathbf{E}^{k+1}-\mathbf{X})(\mathbf{Z}^{k+1})^T\right] \\ \left(\frac{\gamma}{\mu^k}\mathbf{I} + \mathbf{Z}^{k+1}(\mathbf{Z}^{k+1})^T\right)^{-1} \end{aligned} \tag{22}$$

The detailed procedures for solving Eq. (13) are presented in Algorithm 1.

Algorithm 1 : Solving Eq. (13) by Inexact ALM

Input: Training data matrix $\mathbf{X}$; Paramters λ, α, β and γ
Output: $\mathbf{Z}$, $\mathbf{D}$ and $\mathbf{E}$
Initialize: $\mathbf{Z}^0$=0, $\mathbf{J}^0$=0, $\mathbf{L}^0$=0, $\mathbf{E}^0$=0, $\mathbf{Y}_1^0$=0, $\mathbf{Y}_2^0$=0, $\mathbf{Y}_3^0$=0, μ^0=10^{-5}, μ_{max}=10^8, ρ=1.1, ε=10^{-6}
while not converged **do**
Update $\mathbf{J}$ using (14)
Update $\mathbf{Z}$ using (18)
Update $\mathbf{L}$ using (19)
Update $\mathbf{E}$ using (20)
Update $\mathbf{D}$ using (22)
Update the multipliers:
$\mathbf{Y}_1^{k+1} = \mathbf{Y}_1^k + \mu^k(\mathbf{X}-\mathbf{D}^{k+1}\mathbf{Z}^{k+1}-\mathbf{E}^{k+1})$
$\mathbf{Y}_2^{k+1} = \mathbf{Y}_2^k + \mu^k(\mathbf{Z}^{k+1}-\mathbf{J}^{k+1})$
$\mathbf{Y}_3^{k+1} = \mathbf{Y}_3^k + \mu^k(\mathbf{Z}^{k+1}-\mathbf{L}^{k+1})$
Update μ
$\mu^{k+1} = \min(\mu_{max}, \rho\mu^k)$
Check the convergence conditions:
$\|\mathbf{Z}^{k+1}-\mathbf{J}^{k+1}\|_\infty < \varepsilon$, $\|\mathbf{Z}^{k+1}-\mathbf{L}^{k+1}\|_\infty < \varepsilon$ and $\|\mathbf{X}-\mathbf{D}^{k+1}\mathbf{Z}^{k+1}-\mathbf{E}^{k+1}\|_\infty < \varepsilon$



(a) Example images from the Extended Yale B database

(b) Example images from the AR database

(c) Example images from the ORL database

(d) Example images from the LFW database

(e) Example images from the Scene 15 dataset

Fig. 2 Examples of the benchmarking datasets: **a** the Extended Yale B database with illumination variations; **b** the AR database with appearance variations in illumination, expression and occlusion; **c** the ORL database with expression and pose variations; **d** the LFW database including variations in pose, illumination and occlusion; **e** the Scene 15 dataset consisting of various images for a specific scene

5 Experimental results

The proposed RBDS is evaluated on five publicly available databases: Extended Yale B [27], AR [28], ORL [29], LFW [30] and Scene 15 dataset [31]. Example images from these databases are shown in Fig. 2. For the first four face databases, we deal with training and test images being corrupted by factors such as illumination variation, expression changes, pose variation, occlusion and uniformly distributed noise. Similar factors affect the last dataset which is concerned with scene classification. The proposed approach is compared with related low-rank and dictionary learning methods, including BDLRR [25], $D^2L^2R^2$ [13], SLRR [12], DLRD_SR [11], LRSI [16], RPCA [9], FDDL [32], LLC [33], CRC [34], SR [8] and SVM. It should be noted that

SRW indicates the case when all the training data are used as the dictionary. *SRS* has the number of atoms as our proposed RBDS. In order to comprehensively evaluate the role of dictionary learning and the effect of the block-diagonal structure term, our approach (RBDS) is compared with its two special cases: LRRS_BD and LRRS, the objective functions of which are shown in Eqs. (7) and (5), respectively. Each experiment is repeated ten times, and the average recognition results are reported for all approaches.

5.1 Extended Yale B database

The Extended Yale B database has 2414 frontal-face images of 38 subjects. The images of size 192×168 were taken under laboratory-controlled lighting conditions. There are between 59 and 64 images for each person. Following the experimental protocol in RCBD [24], we evaluate our approach on images down sampled by factors 1/2, 1/4 and 1/8, and the resulting feature dimensions are 8064, 2016 and 504, respectively. We randomly select N_c images of each person as the training set ($N_c = 8$, 32), and the remaining ones as the test set. When there are eight training images of each subject, the learned dictionary has five atoms for each class. When $N_c = 32$, the learned dictionary has 20 atoms per class.

The results of different methods obtained on the Extended Yale B database are reported in Table 1. According to the table, our RBDS method achieves the best performance in most cases, even when only a small number of training samples are available. Furthermore, the experimental results indicate that RBDS can handle the challenges of illumination and expression changes. The proposed method outperforms SLRR by 0.7% with 8 training images per person and by 5.2% with 32 training images per person on average. Our approach achieves a significant performance gain in the case of 32 training images per person.

5.2 Evaluation on occluded faces

The AR database consists of over 4000 images of 126 subjects. For each individual, 26 images are taken in different conditions in two separate sessions. There are 13 images from each session, including 3 images with sunglasses, another 3 with scarves and the remaining 7 showing different illumination and expression changes. The resolution of each image is 165×120.

In our experiments, we use a subset of the AR database, which contains 50 male and 50 female subjects. Following the experimental protocol in RCBD [24], we convert the color images into grayscale and down-sample them by a factor of 1/3, resulting in the dimensionality of 2200. We consider the following three scenarios:

(1) *Sunglasses*: We first investigate the effect of occluded samples by sunglasses, which affect about 20% of the face image. We use seven neutral images plus one image with sunglasses (randomly chosen) from Session 1 for training (eight training images per class), and the remaining neutral

Table 2 Recognition accuracy (%) on the AR database

Scenario	Sunglasses	Scarf	Mixed
RBDS	**95.5**	**93.3**	**93.7**
LRRS_BD	90.6	86.8	85.7
LRRS	89.2	85.2	85.6
BDLRR [25]	90.6	88.4	87.8
$D^2L^2R^2$ [13]	89.3	84.1	82.7
SLRR [12]	87.3	83.4	82.4
LRSI [16]	84.9	76.4	80.3
RPCA [9]	83.2	75.8	78.9
SRW [8]	86.8	83.2	79.2
SRS [8]	82.1	72.6	65.5
LLC [33]	65.3	59.2	59.9

Bold values indicate the best recognition accuracy

Table 1 Recognition accuracy (%) on the extended Yale B database

No. per class	$N_c = 8$			$N_c = 32$		
Sample rate	1/8	1/4	1/2	1/8	1/4	1/2
RBDS	**80.3**	82.8	83.3	**97.2**	**98.7**	**98.9**
LRRS_BD	76.0	79.7	81.1	96.7	97.9	98.0
LRRS	75.3	78.6	79.5	96.8	96.9	97.7
BDLRR [25]	73.4	75.1	77.6	95.8	96.8	97.5
$D^2L^2R^2$ [13]	79.3	82.8	**84.0**	96.1	97.2	97.5
SLRR [12]	76.6	**83.7**	83.8	89.9	93.6	95.7
LRSI [16]	73.3	80.9	80.8	89.5	93.5	94.5
RPCA [9]	74.6	78.3	80.2	85.6	90.7	94.1
SRW [8]	79.3	83.0	83.8	87.2	89.5	90.7
SRS [8]	75.3	78.9	80.1	84.4	85.7	85.9
LLC [33]	65.7	70.6	76.1	76.4	80.0	85.6

Bold values indicate the best recognition accuracy

images (all from Session 2) and the rest of the images with sunglasses (two taken from Session 1 and three from Session 2) for testing (12 test images per class).

(2) *Scarf*: Here we replace the images with sunglasses in the above scenario by images with a scarf.

(3) *Mixed (Sunglasses + Scarf)*: In the last scenario, the training samples may be occluded by either sunglasses or scarves, which is more challenging than the above two scenarios. Seven neutral images and two corrupted images (one with sunglasses and one with scarf) from Session 1 are used for training (nine training images per class), and the remaining ones are used for testing (17 test images per class).

Similar to RCBD [24], a compact dictionary with five atoms per class is learned under different scenarios. Table 2 summarizes the experimental results on the AR database. Our approach consistently performs the best and its accuracy gains over BDLRR are 4.9% for the sunglasses scenario, 4.9% for the scarf scenario and 5.9% for the mixed scenario, respectively. As expected, LRRS_BD is inferior to RBDS, which demonstrates that a high-quality dictionary is needed for learning a discriminative representation when both training and test images are corrupted.

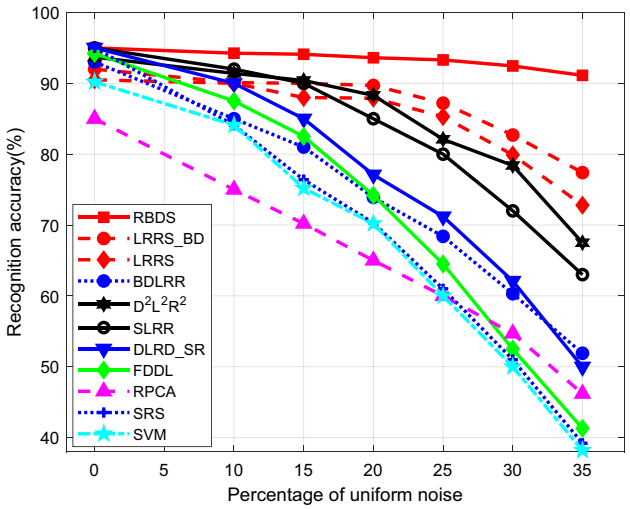


Fig. 3 Recognition accuracy on the AR database with different levels of pixel noise

5.3 Evaluation on pixel corruption

In this section, we evaluate the proposed method on the AR database with different levels of corruption. First, we select seven neutral images with illumination and expression changes from Session 1 for training and the other seven neutral images from Session 2 for testing. A certain percentage of randomly selected pixels from both training and test images are replaced by noise uniformly distributed between the minimal and maximal pixel value. The number of dictionary atoms per class is set to 7. The recognition accuracy is plotted under different levels of corruption in Fig. 3. Our approach outperforms $D^2L^2R^2$ by 8.8% on average. Figure 3 demonstrates that the proposed RBDS consistently outperforms all the other approaches for all levels of pixel corruption.

5.4 Evaluation on block occlusion

To further verify the performance of different methods in tackling random block occlusion, an experiment is carried out on the ORL database. This database has 400 images of 40 individuals. The images were taken at different times, with lighting variation, facial expression and pose changes. We crop and normalize each image to 28×23 pixels. For each subject, a half of the images are randomly selected as training samples, and the remaining ones serve as test samples. We replace a randomly located block of each image with an unrelated random image. The experiments are conducted for different degrees of block occlusion.

Table 3 presents the recognition accuracy for different levels of occlusion on the ORL database. Our approach (RBDS) achieves the best performance and outperforms DLRD_SR by 1.9% on average. Our approach shows high

Table 3 Recognition accuracy (%) on the ORL database with block occlusion

Noise percent	0	10	20	30	40	50
RBDS	96.0	**94.7**	**91.9**	**89.2**	**80.1**	**73.5**
LRRS_BD	96.1	94.7	91.1	86.0	79.9	70.6
LRRS	96.1	93.6	89.8	84.8	77.7	70.9
BDLRR [25]	95.3	91.3	83.2	72.2	55.6	47.0
$D^2L^2R^2$ [13]	94.5	94.0	89.5	86.5	76.0	70.5
DLRD_SR [11]	95.9	94.4	91.1	86.0	76.7	69.9
FDDL [32]	**96.7**	94.0	89.8	85.1	76.1	68.3
RPCA [9]	89.3	88.0	83.0	76.6	72.0	66.2
SRC [8]	95.2	91.7	86.0	75.8	61.8	54.0
SVM [11]	94.6	88.5	80.6	71.6	57.3	42.0

Bold values indicate the best recognition accuracy

Table 4 Recognition accuracy (%) on the LFW-a dataset

Methods	Accuracy
RBDS	**71.7**
LRRS	62.1
BDLRR [25]	68.6
$D^2L^2R^2$ [13]	68.6
SLRR [12]	68.2
LRSI [16]	66.2
RPCA [9]	66.3
CRC [34]	64.6
SRC [8]	68.3
LLC [33]	60.1
SVM	58.2

Bold values indicate the best recognition accuracy

Table 5 Recognition accuracy (%) achieved on the Scene 15 dataset

Methods	Accuracy
RBDS	**98.66**
LRRS	95.54
BDLRR [25]	98.50
$D^2L^2R^2$ [13]	96.58
SLRR [12]	92.90
LRSI [16]	92.46
RPCA [9]	89.10
CRC [34]	92.00
SRC [8]	91.80
LLC [33]	89.20
SVM	95.06

Bold values indicate the best recognition accuracy

robustness to the severe corruption posed by block occlusion and achieves 3.4% improvement in the case of 40% block noise. Moreover, thanks to the block-diagonal structure term, LRRS_BD achieves better results than LRRS. However, LRRS_BD is inferior to our proposed RBDS since it does not involve the dictionary learning process.

5.5 Unconstrained face classification

Thus far, we have conducted the experiments on constrained datasets which do not exhibit large appearance variations of the same identity. In unconstrained scenarios face images of the same subject change dramatically due to variation in pose, illumination, expression and occlusion. Furthermore, test images may not contain the same type of variations and occlusions as the training images. To evaluate the robustness of our method in unconstrained scenarios, we conduct experiments on the LFW database. This dataset contains images of 5,749 individuals and we use the LFW-a, which is an aligned version of LFW obtained using a commercial face alignment software. We select the subjects that include no less than ten samples and we construct a dataset with 158 subjects from LFW-a. For each person, we randomly select five samples for training (resulting in a dictionary of 790 faces) and the other five for testing. The images are resized to 90×90. The experimental results are shown in Table 4, where we can observe that our approach is superior to the competing methods. This again testifies to the effectiveness of RBDS.

5.6 Scene categorization

The last experiment is performed on the Scene 15 dataset [31]. This dataset contains 4485 images in total of 15 categories of natural scenes. Each class has 200 to 400 images, and the average image size is about 250×300 pixels. This database consists of a variety of outdoor and indoor scenes, such as office, kitchen, tall building and country scenes. The 3000-dimensional SIFT-based features provided in [35] are exploited in our experiments. Following the common experimental setting used in [31] and [35], 100 images per class are randomly selected as training data and the remaining images are used for testing. The comparative results of all the approaches are presented in Table 5. It can be seen that the proposed RBDS has the best performance. Note that compared with the recently proposed BDLRR method, RBDS achieves a modest 0.16% improvement.

6 Conclusion

In this paper, we presented a low-rank-based method to learn image representations promoted by a block-diagonal structure constraint, i.e., RBDS, for pattern classification. A regularization term is incorporated into the framework of LRR to capture structure information globally. With this term, the off-block-diagonal elements of the representation matrix are minimized. As a result, the correlations between distinct classes are reduced while the coherence of intraclass representation is boosted. A compact dictionary is learned as part of the training process. We also proposed an effective algorithm to solve the optimization problem defined by our novel formulation. The experimental results obtained on five public datasets show that RBDS offers better recognition performance on average, and it is robust to appearance variations in illumination, expression, occlusion and random pixel corruption.

Acknowledgements The work was supported by the National Natural Science Foundation of China (61672265, U1836218, 61902153, 61876072), the 111 Project of the Ministry of Education of China (B12018), the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant No. KYLX_1123, the Overseas

Studies Program for Postgraduates of Jiangnan University and the China Scholarship Council (CSC, No.201706790096), the EPSRC programme grant (FACER2VM) under the number EP/N007743/1, the U.S. Army Research Laboratory, the U. S. Army Research Office, the U.K. Ministry of Defence and the U.K. EPSRC grant under the number EP/R013616/1.

References

1. Zheng J, Qiu H, Sheng W, Yang X, Yu H (2018) Kernel group sparse representation classifier via structural and non-convex constraints. Neurocomputing 296:1–11
2. Shao C, Song X, Feng ZH et al (2017) Dynamic dictionary optimization for sparse-representation-based face classification using local difference images. Inf Sci 393:1–14
3. Liu G (2018) Robust visual tracking via smooth manifold kernel sparse learning. IEEE Trans Multimed 20(11):2949–2963
4. Wang J, Shi D, Cheng D, Zhang Y, Gao J (2016) LRSR: low-rank-sparse representation for subspace clustering. Neurocomputing 214:1026–1037
5. Song X, Feng Z, Hu G et al (2018) Dictionary integration using 3D morphable face models for pose-invariant collaborative-representation-based classification. IEEE Trans Inf Forensics Secur 13(11):2734–2745
6. Chhatrala R, Patil S, Lahudkar S, Jadhav D (2019) Sparse multilinear Laplacian discriminant analysis for gait recognition. Pattern Anal Appl 22(2):505–518
7. Song X, Feng Z, Hu G et al (2017) Half-face dictionary integration for representation-based classification. IEEE Trans Cybern 47(1):142–152
8. Wright J, Yang A, Ganesh A et al (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
9. Candès E, Li X, Ma Y, Wright J (2011) Robust principal component analysis? J ACM 58(3):11
10. Liu G, Lin Z, Yan S et al (2013) Robust recovery of subspace structures by low-rank representation. IEEE Trans Pattern Anal Mach Intell 35(1):171–184
11. Ma L, Wang C, Xiao B, Zhou W (2012) Sparse representation for face recognition based on discriminative low-rank dictionary learning. In: 2012 IEEE conference on computer vision and pattern recognition, pp 2586–2593
12. Zhang Y, Jiang Z, Davis L (2013) Learning structured low-rank representations for image classification. In: 2013 IEEE conference on computer vision and pattern recognition, pp 676–683
13. Li L, Li S, Fu Y (2014) Learning low-rank and discriminative dictionary for image classification. Image Vis Comput 32(10):814–823
14. Nguyen H, Yang W, Sheng B, Sun C (2016) Discriminative low-rank dictionary learning for face recognition. Neurocomputing 173:541–551
15. Zheng Z, Yu M, Jia J et al (2014) Fisher discrimination based low rank matrix recovery for face recognition. Pattern Recogn 47(11):3502–3511
16. Wei C, Chen C, Wang Y (2014) Robust face recognition with structurally incoherent low-rank matrix decomposition. IEEE Trans Image Process 23(8):3294–3307
17. Yin H, Wu X (2016) Face recognition based on structural incoherence and low rank projection. In: 2016 International conference on intelligent data engineering and automated learning, pp 68–78
18. Chen J, Zhang Y (2014) Sparse representation for face recognition by discriminative low-rank matrix recovery. J Vis Commun Image Represent 25(5):763–773
19. Dong Z, Pei M, Jia Y (2016) Orthonormal dictionary learning and its application to face recognition. Image Vis Comput 51:13–21
20. Rong Y, Xiong S, Gao Y (2017) Low-rank double dictionary learning from corrupted data for robust image classification. Pattern Recogn 72:419–432
21. Gao G, Yang J, Jing XY et al (2017) Learning robust and discriminative low-rank representations for face recognition with occlusion. Pattern Recogn 66:129–143
22. Du H, Zhao Z, Wang S, Zhang F (2018) Discriminative low-rank graph preserving dictionary learning with Schatten-p quasi-norm regularization for image recognition. Neurocomputing 275:697–710
23. Wu C, Ding J (2018) Occluded face recognition using low-rank regression with generalized gradient direction. Pattern Recogn 80:256–268
24. Li Y, Liu J, Lu H, Ma S (2014) Learning robust face representation with classwise block-diagonal structure. IEEE Trans Inf Forensics Secur 9(12):2051–2062
25. Zhang Z, Xu Y, Shao L, Yang J (2018) Discriminative block-diagonal representation learning for image recognition. IEEE Trans Neural Netw Learn Syst 29(7):3111–3125
26. Lin Z, Liu R, Su Z (2011) Linearized alternating direction method with adaptive penalty for low-rank representation. In: 2011 advances in neural information processing systems, pp 612–620
27. Georghiades A, Belhumeur P, Kriegman D (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans Pattern Anal Mach Intell 23(6):643–660
28. Martinez AM (1998) The AR face database. CVC technical report, p 24
29. Samaria F, Harter A (1994) Parameterisation of a stochastic model for human face identification. In: Proceedings of the second IEEE workshop on applications of computer vision, pp 138–142
30. Huang G, Mattar M, Berg T et al (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report 07–49. University of Massachusetts, Amherst
31. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: 2006 IEEE conference on computer vision and pattern recognition, pp 2169–2178
32. Yang M, Zhang L, Feng X, Zhang D (2011) Fisher discrimination dictionary learning for sparse representation. In: 2011 IEEE international conference on computer vision, pp 543–550
33. Wang J, Yang J, Yu K, Lv F, Huang T (2010) Locality-constrained linear coding for image classification. In: 2010 IEEE conference on computer vision and pattern recognition, pp 3360–3367
34. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: which helps face recognition? In: 2011 IEEE international conference on computer vision, pp 471–478
35. Jiang Z, Lin Z, Davis L (2013) Label consistent K-SVD: learning a discriminative dictionary for recognition. IEEE Trans Pattern Anal Mach Intell 35(11):2651–2664