sensors

Article

Block-Diagonal Constrained Low-Rank and Sparse Graph for Discriminant Analysis of Image Data

Tan Guo *, Xiaoheng Tan *, Lei Zhang *, Chaochen Xie and Lu Deng

College of Communication Engineering, Chongqing University, Chongqing 400044, China;
xie_cc1@cqu.edu.cn (C.X.); ludeng@cqu.edu.cn (L.D.)
* Correspondence: tanguo@cqu.edu.cn (T.G.); txh@cqu.edu.cn (X.T.); leizhang@cqu.edu.cn (L.Z.)

Received: 29 March 2017; Accepted: 19 June 2017; Published: 22 June 2017

Abstract: Recently, low-rank and sparse model-based dimensionality reduction (DR) methods have aroused lots of interest. In this paper, we propose an effective supervised DR technique named block-diagonal constrained low-rank and sparse-based embedding (BLSE). BLSE has two steps, i.e., block-diagonal constrained low-rank and sparse representation (BLSR) and block-diagonal constrained low-rank and sparse graph embedding (BLSGE). Firstly, the BLSR model is developed to reveal the intrinsic intra-class and inter-class adjacent relationships as well as the local neighborhood relations and global structure of data. Particularly, there are mainly three items considered in BLSR. First, a sparse constraint is required to discover the local data structure. Second, a low-rank criterion is incorporated to capture the global structure in data. Third, a block-diagonal regularization is imposed on the representation to promote discrimination between different classes. Based on BLSR, informative and discriminative intra-class and inter-class graphs are constructed. With the graphs, BLSGE seeks a low-dimensional embedding subspace by simultaneously minimizing the intra-class scatter and maximizing the inter-class scatter. Experiments on public benchmark face and object image datasets demonstrate the effectiveness of the proposed approach.

Keywords: dimensionality reduction; low-rank representation; sparse representation; block-diagonal; image classification

1. Introduction

With the rapid development of information technology, nowadays high precision sensors sense large-scale data, especially image data, all the time. These data often feature high dimensionality, and consist of redundant information or noise. How to analyze these high-dimensional data has attracted the interest of many researchers. Dimensionality reduction (DR) is a practical way to deal with this problem. DR aims to find a lower-dimensional embedding subspace where some desired properties can be preserved as much as possible [1–3]. The most well-known DR methods are, for example, principal component analysis (PCA) [4] and linear discriminate analysis (LDA) [5]. PCA is unsupervised, and applies orthogonal projection to maximize data variance. As a supervised method, LDA finds the linear projection axes on which the ratio between the between-class and within-class scatters is maximized. LDA cannot be directly applied to small size sample (SSS) problem because the within-class scatter matrix is singular. To avoid this problem, Li et al. [6] adopted the difference of between-class scatter and within-class scatter as the discriminating criterion for embedding learning. The method, termed maximum margin criterion (MMC), is simple and effective. However, these linear methods cannot reveal the essential structure of data with non-linear distributions. With kernel tricks, kernel principal component analysis (KPCA) [7] and kernel Fisher discriminate analysis (KFD) [8] were developed to handle data with non-linearity structures. Some manifold learning algorithms such as LLE [9], Isomap [10], and Laplacian eigenmaps (LE) [11] have also been presented to discover the intrinsic manifold structure in data.

Yan et al. [12] proposed a general DR framework called graph embedding (GE), where some DR methods, e.g., PCA, LDA, Isomap, LLE, and LE, could be considered as special instances under the framework. Based on GE, some discriminant DR methods, e.g., marginal Fisher analysis (MFA) [12], neighborhood preserving discriminate embedding (NPDE) [13] and sparse discriminate manifold embedding (SDME) [14], have been developed. The differences between these methods lie in the design of intrinsic and penalty graphs and the type of embedding. Graph construction has become the key of most GE-based DR methods. However, the way to establish high-quality graphs is an open problem. Common graph construction methods include k-nearest neighbor and ε-radius ball, both of which connect graph vertices with simple rules which are, however, highly sensitive to dataset noise and it is difficult to determine the parameters for real-world applications.

Sparse and low-rank models have been widely applied for visual analysis. Sparse representation (SR) and low rank representation (LRR) are utilized to construct the affinity matrix (or graph) [15–19]. They express each datum as a linear combination of all other data points, and use the representation coefficient to measure the similarity or neighborhood relationship of samples. Contrastively, sparse graphs are able to preserve local linear structure but lack global constraints. LRR takes the correlation structure of data into account, and finds a low-rank representation instead of a sparse one. The low-rank property has been proved to be effective in preserving global data structures.

The block-diagonal structure is often desired in affinity matrix construction. Ideally, the affinity matrix should be block-diagonal, and the inter-class affinities are all zeros [20]. However, SR and LRR can only generate a block-diagonal affinity matrix under restrictive conditions. For example, it has been shown that when the subspaces are independent, the solution to LRR is block-diagonal [20]. For a block-diagonal structured solution, Zhao et al. [21] incorporated least square regression and graph regularization to construct a sparse graph with block-wise constraint (SGB) for face representation. The method can well uncover the global structure of the multiple subsets of the data, and preserve the local intrinsic information. The performance can be further boosted by introducing the idea of ensemble learning [22]. Tang et al. [23] developed an extension of LRR termed as structure-constrained LRR (SC-LRR) for general disjoint subspace clustering by introducing an explicit structure constraint. Zhang et al. [24] presented a discriminative, structured low-rank method to explore structure information by incorporating an ideal-code regularization term. Li et al. [25] proposed to restrict the representation coefficients outside the diagonal block to be small to get a block-diagonal structure representation. Alternatively, Feng et al. [26] explicitly pursued a block-diagonal structure solution via a graph Laplacian constraint-based formulation. In [27], sparse graph-based discriminate analysis (SGDA) was developed by preserving the sparse connection in a block-structured affinity matrix with class-specific samples. Similarly, low-rank graph-based discriminate analysis (LGDA) [28] preserves the global structure in data using low-rank constraints. Meanwhile, low-rank and sparse graphs have been applied for semi-supervised learning [29,30]. A sparse and low-rank graph-based discriminate analysis (SLGDA) method was developed in [28] to purse block-diagonal structured affinity matrix with both sparsity and low-rank constraints. However, SGDA, LGDA and SLGDA consider intra-class affinity relationship class by class, which suffers from high computation costs. Besides, these methods find the representation of each sample using only the intra-class samples, which might not be able to reveal inter-class adjacent relationships.

In this paper, motivated by recent achievements in SR and LRR [15–30], we propose a novel supervised DR method, namely block-diagonal constrained low-rank and sparse-based embedding (BLSE). The model has two steps: first, a self-expressive model, i.e., block-diagonal constrained low-rank and sparse representation (BLSR) model is developed to reveal the intra-class and inter-class adjacent relationships among samples and discover the local and global structures latent in data. The desired representation learning example is shown in Figure 1. Here, different colors stand for samples of different classes. With the block-diagonal constraint, each sample is encouraged to be represented by intra-class samples. The obtained representation matrix $\mathbf{Z}$ tends to be block-diagonal and has good identification capability highlighting both intra-class similarities and inter-class differences. Due to these merits, the intra-class and inter-class representations obtained by BLSR are utilized to construct corresponding intra-class and inter-class graphs.

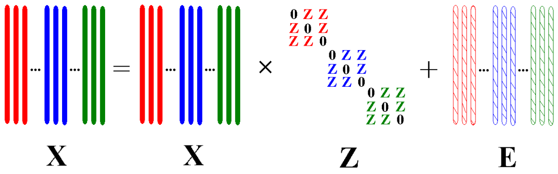


Figure 1. An example for desired block-diagonal constrained low-rank and sparse representation (BLSR). Samples in dataset $\mathbf{X}$ are encouraged to be represented by samples from the same class with noise $\mathbf{E}$ removed, and the representation matrix $\mathbf{Z}$ tends to have a block-diagonal structure.

Then, as shown in Figure 2, the block-diagonal constrained low-rank and sparse graph embedding (BLSGE) method finds a low-dimensional subspace with enhanced intra-class compactness and inter-class separation using the graphs. It is worth noting that there are some major differences between our BLSR model and the model presented in [23], although they have similar formulations. In that method, a weight matrix is defined to provide a moderate amount of correct information for the solution. Different from their strategy, we separately optimize the solution to be sparse, and develop an iteration method to explicitly optimize the diagonal elements of the solution to be large, and the rest ones to be small via a predefined block-diagonal mask matrix. Since our aim is to induce inter-class and inter-class graphs from the solution for further embedding learning, our strategy of promoting intra-class affinity weights large and inter-class affinity weights small is more applicable for the problem of interest. In sum, the main contributions of this paper are as follows:

(1) A self-expressive model, i.e., BLSR is devised by incorporating sparsity, low rankness as well as a novel block-diagonal constraint. BLSR can not only simultaneously capture the local and global structures, but also highlight both the intra-class similarities and inter-class differences of samples.
(2) With the intra-class and inter-class graphs derived from BLSR, BLSGE seeks an optimal feature space by simultaneously minimizing the intra-class scatter and maximizing the inter-class scatter. Generally, a novel supervised dimensionality reduction method namely BLSE is developed by taking the advantages of BLSR and GE framework.
(3) BLSE is applied for the dimensionality reduction and classification of visual data. Extensive experiments on the public face and object datasets verify the effective of proposed method.

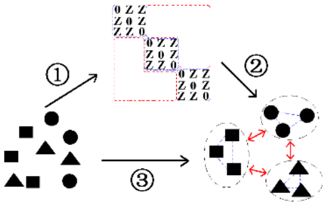


Figure 2. Illustration for block-diagonal constrained low-rank and sparse based embedding (BLSE) model. ① BLSR is applied to get the block-diagonal constrained low-rank and sparse representation of data. ② The representation results of BLSR is utilized to construct the intra-class and inter-class graphs. ③ BLSGE finds a low dimensional embedding with enhanced intra-class compactness and inter-class separation using the graphs.

The remainder of this paper is organized as follows: in Section 2, we briefly introduce some related works. In Section 3, we will introduce our BLSE model. Its two steps, i.e., BLSR and BLSGE, will be presented in detail. The experimental results are given in Section 4. Finally, we provide the discussion and conclusions in Section 5.

2. Related Works

Let us suppose a labeled dataset $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \Re^{\mathrm{D} \times n}$, D is the dimension of sample in original space. n is the number of training samples with n_k $(k = 1,2,\dots,C)$ samples per class. The label of $\mathbf{x}_i (i = 1,2,\dots,n)$ is denoted as $l(\mathbf{x}_i)$. Data points in $\mathbf{X}$ are ordered, as is common, in terms of their class labels. $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] \in \Re^{d \times n}$ (typically d << D) is the lower dimensional projected data of $\mathbf{X}$ with projection matrix $\mathbf{V} \in \Re^{D \times d}$.

2.1. Low Rank and Sparse Representation

Low rank and sparse models have been seen a surge of interests in recent years, and been successfully exploited in many applications, such as subspace clustering [16,17], face recognition [25,31,32], head pose estimation [33], information processing [34–37], transfer learning [38,39], and extreme learning machine [40]. Low-rankness is an appropriate criterion to capture low-rank dimensional structure in high-dimensional data, and low-rank representation (LRR) is robust to sparse noise. Sparse representation (SR) has been shown good discrimination capacity. Low-rank and sparse representation is pursued to take the merits of both the two aspects. Learning the low-rank and sparse representation $\mathbf{Z}$ of dataset $\mathbf{X}$ on dictionary $\mathbf{D}$ can be formulated as follows:

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

where $\|\cdot\|_l$ is used to characterize noise $\mathbf{E}$. It can be sparse noise $\|\mathbf{E}\|_0$ or sample specific noise $\|\mathbf{E}\|_{2,1}$. λ and β control the effects of noise term $\mathbf{E}$ and sparse representation term $\mathbf{Z}$. Since the l_0-norm and rank minimization problems are non-convex, the problem is NP-hard. Alternatively, rank function can be relaxed with nuclear norm, which is defined as the sum of the singular values of a matrix. The l_0-norm can be surrogated with l_1-norm. Thus, one can get the following relaxed optimization problem:

$$\min_{\mathbf{Z},\mathbf{E}} \ \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_l + \beta \ \|\mathbf{Z}\|_1 \ \text{ s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \tag{2}$$

With dataset $\mathbf{X}$ itself as the dictionary, Reference [28] proposed the following optimization problem with sample-specific noise:

$$\min_{\mathbf{Z},\mathbf{E}} \ \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} + \beta \ \|\mathbf{Z}\|_1 \ \text{ s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \mathrm{diag}(\mathbf{Z}) = \mathbf{0} \tag{3}$$

where $\mathrm{diag}\ (\mathbf{Z})$ represents the vector containing the diagonal elements of $\mathbf{Z}$, and $\mathbf{0}$ is a zero vector. The obtained low rank and sparse representation matrix $\mathbf{Z}$ can be utilized to construct intrinsic graph for DR [28].

2.2. Graph Embedding

The GE framework provides a unified perspective to understand many DR algorithms [12]. In GE, an intrinsic graph $\mathbf{G} = \{\mathbf{X}, \mathbf{W}\}$ that describes certain desired statistical or geometrical properties of data, and a penalty graph $\mathbf{G}^{\mathrm{P}} = \{\mathbf{X}, \mathbf{W}^{\mathrm{P}}\}$ characterizes a statistical or geometric property which should be avoided need to be constructed. Both $\mathbf{G}$ and $\mathbf{G}^{\mathrm{P}}$ are undirected weighted graphs. $\mathbf{X}$ is the vertex set. $\mathbf{W} \in \Re^{n \times n}$ and $\mathbf{W}^{\mathrm{p}} \in \Re^{n \times n}$ are the weight matrices.

Assuming that the low-dimensional vector representations of the vertices can be obtained from a linear projection as $\mathbf{y} = \mathbf{V}^{\mathrm{T}}\mathbf{x}$. The purpose of GE is to map each vertex of graph into a low-dimensional space that preserves the similarity between the vertex pairs. Then an optimal low-dimensional embedding is given by the graph preserving criterion as:

$$\mathbf{V}^* = \mathop{\arg\min}_{\mathbf{V}^{\mathrm{T}}\mathbf{X}\mathbf{L}_{\mathrm{p}}\mathbf{X}^{\mathrm{T}}\mathbf{V}=\mathbf{I}} \sum_{i \neq j} \left\|\mathbf{V}^{\mathrm{T}}\mathbf{x}_i - \mathbf{V}^{\mathrm{T}}\mathbf{x}_j\right\|^2 \mathbf{W}_{ij} \tag{4}$$

$$= \underset{V^T\mathbf{X}\mathbf{L}_\mathrm{p}\mathbf{X}^\mathrm{T}\mathbf{V}=\mathbf{I}}{\arg\min}\ \mathrm{tr}(\mathbf{V}^\mathrm{T}\mathbf{X}\mathbf{L}\mathbf{X}^\mathrm{T}\mathbf{V})$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix. $\mathbf{D}$ is a diagonal matrix with $\mathbf{D}_{ii} = \sum_{j=1}^{\mathrm{N}} \mathbf{W}_{ij}$. The weight $\mathbf{W}_{ij}$ is used to measure the similarity of the edge connecting vertices. $\mathbf{L}^\mathrm{P}$ is the Laplacian matrix of the penalty graph $\mathbf{G}^\mathrm{P}$ or a simple scale normalization constraint. The linearization extension of graph embedding is computationally efficient for both projection learning and final classification. The construction of intrinsic graph and penalty graph becomes the crux of most dimensionality reduction methods. Besides, the intrinsic and penalty graphs could be, as our method shows, the intra-class and inter-class graphs.

3. Proposed Method

In Section 3.1, we will detail the two steps of BLSE, i.e., BLSR and BLSGE. The optimization processes for BLSR and BLSGE will be given in Section 3.2. Section 3.3 describes the classification process.

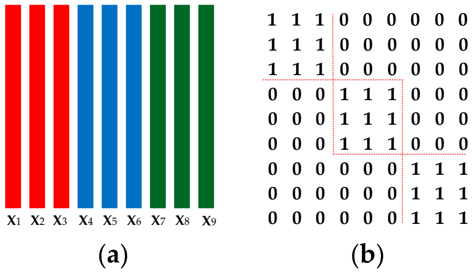
3.1. Block-Diagonal Constrained Low-Rank and Sparse Based Embedding (BLSE)

3.1.1. Block-Diagonal Constrained Low–Rank and Sparse Representation (BLSR)

To reveal the intra-class and inter-class adjacent relationships and discover the local and global structures in data, a self-expressive model, i.e., BLSR is firstly developed. The label information of data is harnessed by introducing a block-diagonal constraint to purse a block-diagonal solution. Specifically, the BLSR model is formulated as:

$$\min_{\mathbf{Z},\mathbf{E}} \ \|\mathbf{Z}\|_* + \frac{\alpha}{2}\|\mathbf{Z} - \mathbf{Z} \odot \mathbf{M}\|_F^2 + \beta\,\|\mathbf{Z}\|_1 + \gamma\|\mathbf{E}\|_1 \ \text{ s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \mathrm{diag}(\mathbf{Z}) = \mathbf{0} \tag{5}$$

where α, β and γ are the trade-off parameters for each component, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. In (5), we try to discover the block-diagonal structure of the resolution via the block-diagonal regularization $\|\mathbf{Z} - \mathbf{Z} \odot \mathbf{M}\|_F^2$, where $\odot$ is the Hadamard product operator of matrices and $\mathbf{M}$ is a predefined mask matrix with an ideal block-diagonal structure. Figure 3 shows an example for the definition of $\mathbf{M}$. $\mathbf{Z} \odot \mathbf{M}$ is used to extract the intra-class representation coefficients for each sample. By minimizing $\|\mathbf{Z} - \mathbf{Z} \odot \mathbf{M}\|_F^2$, representation coefficient of each sample corresponding to the inter-class samples is promoted to be small, but not necessarily be zero. Each sample is encouraged to be represented by the intra-class samples. The obtained block-diagonal representation matrix $\mathbf{Z}$ has good identification capability highlighting both the intra-class similarities and inter-class differences.



$$\begin{array}{ccc:ccc:ccc}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hdashline
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
\hdashline
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{array}$$

Figure 3. An illustration for the definition of matrix $\mathbf{M}$. (**a**) 9 training samples from 3 classes with 3 samples per class; (**b**) The defined mask matrix $\mathbf{M}$. If $l(\mathbf{x}_i) = l(\mathbf{x}_j)$, $\mathbf{M}_{i,j} = 1$, $\mathbf{M}_{j,i} = 1$, otherwise $\mathbf{M}_{i,j} = 0$, $\mathbf{M}_{j,i} = 0$. The diagonal-block elements of $\mathbf{M}$ are all ones, and the rest are zeros.

3.1.2. Block-Diagonal Constrained Low–Rank and Sparse Graph Embedding (BLSGE)

The representation matrix $\mathbf{Z}$ of BLSR is then employed as affinity weights matrix to construct intra-class graph $\mathbf{G}^{\mathrm{intra}} = \{\mathbf{X}, \mathbf{W}^{\mathrm{intra}}\}$ and inter-class graph $\mathbf{G}^{\mathrm{inter}} = \{\mathbf{X}, \mathbf{W}^{\mathrm{inter}}\}$. First, we define the affinity matrix as $\mathbf{W} = (|\mathbf{Z}| + |\mathbf{Z}^{\mathrm{T}}|)/2$, then the connecting weights for intra-class and inter-class graphs are respectively defined as:

$$\mathbf{W}_{ij}^{\text{intra}} = \begin{cases} \mathbf{W}_{ij}, \text{if } \mathbf{W}_{ij} \neq 0 \text{ and } l(\mathbf{x}_{\mathrm{i}}) = l(\mathbf{x}_j) \\ 0, \text{otherwise} \end{cases} \tag{6}$$

$$\mathbf{W}_{ij}^{\text{inter}} = \begin{cases} \mathbf{W}_{ij}, \text{if } \mathbf{W}_{ij} \neq 0 \text{ and } l(\mathbf{x}_i) \neq l(\mathbf{x}_j) \\ 0, \text{otherwise} \end{cases} \tag{7}$$

Whether two arbitrary points in the graphs are connected or not and the connection weight are adaptively determined by our BLSR model. It is desired that samples from the same class in feature space should be as close as possible, and those from different classes should be as far as possible. With projection matrix $\mathbf{V}$, the optimization objective functions are defined as:

$$\begin{cases} \min_{V} \frac{1}{2} \sum_{i,j} \left\| \mathbf{V}^{\mathrm{T}} \mathbf{x}_i - \mathbf{V}^{\mathrm{T}} \mathbf{x}_j \right\|^2 \mathbf{W}_{ij}^{\text{intra}} \\ \max_{V} \frac{1}{2} \sum_{i,j} \left\| \mathbf{V}^{\mathrm{T}} \mathbf{x}_i - \mathbf{V}^{\mathrm{T}} \mathbf{x}_j \right\|^2 \mathbf{W}_{ij}^{\text{inter}} \end{cases} \tag{8}$$

For ease of classification, a big $\mathbf{W}_{ij}^{\text{intra}}$ is required in (8) to make the projected samples from the same class close to each other, and a small $\mathbf{W}_{ij}^{\text{inter}}$ is needed in (8) to make the projected samples from different classes far away from each other. The requirements can be satisfied by the representation strategy in BLSR. With some mathematical operations, we have:

$$\begin{cases} \min_{V} \operatorname{tr}(\mathbf{V}^{\mathrm{T}} \mathbf{X} \mathbf{L}^{\text{intra}} \mathbf{X}^{\mathrm{T}} \mathbf{V}) \\ \max_{V} \operatorname{tr}(\mathbf{V}^{\mathrm{T}} \mathbf{X} \mathbf{L}^{\text{inter}} \mathbf{X}^{\mathrm{T}} \mathbf{V}) \end{cases} \tag{9}$$

Then, the objective function of BLSGE can be formulated as:

$$\min_{V} \frac{\operatorname{tr}(\mathbf{V}^{\mathrm{T}} \mathbf{X} \mathbf{L}^{\text{intra}} \mathbf{X}^{\mathrm{T}} \mathbf{V})}{\operatorname{tr}(\mathbf{V}^{\mathrm{T}} \mathbf{X} \mathbf{L}^{\text{inter}} \mathbf{X}^{\mathrm{T}} \mathbf{V})} \tag{10}$$

3.2. Optimizations for BLSR and BLSGE

3.2.1. Optimization for BLSR

We convert problem (5) into the following equivalent problem by introducing auxiliary variables $\mathbf{J}$ and $\mathbf{L}$:

$$\min_{\mathbf{Z},\mathbf{E},\mathbf{J},\mathbf{L}} \|\mathbf{J}\|_* + \frac{\alpha}{2} \|\mathbf{Z} - \mathbf{Z} \odot \mathbf{M}\|_{\mathrm{F}}^2 + \beta \|\mathbf{L}\|_1 + \lambda \|\mathbf{E}\|_1 \ \text{ s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \mathbf{Z} = \mathbf{J}, \mathbf{Z} = \mathbf{L} \tag{11}$$

We then have the corresponding Augmented Lagrange Multipliers (ALM) [41] function:

$$\begin{aligned} \Xi = \min_{\mathbf{Z},\mathbf{E},\mathbf{J},\mathbf{L}} & \|\mathbf{J}\|_* + \frac{\alpha}{2} \|\mathbf{Z} - \mathbf{Z} \odot \mathbf{M}\|_F^2 + \beta \|\mathbf{L}\|_1 + \lambda \|\mathbf{E}\|_1 + \langle \mathbf{Y}_1, \mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{E} \rangle \\ & + \langle \mathbf{Y}_2, \mathbf{Z} - \mathbf{J} \rangle + \langle \mathbf{Y}_3, \mathbf{Z} - \mathbf{L} \rangle + \frac{\mu}{2} (\|\mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{E}\|_F^2 + \|\mathbf{Z} - \mathbf{J}\|_F^2 + \|\mathbf{Z} - \mathbf{L}\|_F^2) \end{aligned} \tag{12}$$

where $\mathbf{Y}_1 \in \Re^{D \times n}$, $\mathbf{Y}_2 \in \Re^{n \times n}$ and $\mathbf{Y}_3 \in \Re^{n \times n}$ are Lagrange multipliers and $\mu > 0$ is a penalty parameter. The problem can be solved iteratively by updating each variable with others fixed. The steps to solve the problem in $(k + 1)$-th iteration are as follows:

*Step 1 (*Update $\mathbf{Z}$): $\mathbf{Z}$ can be updated by solving the following optimization problem (13):

$$\begin{aligned} \mathbf{Z}^{k+1} = \underset{Z}{\operatorname{argmin}} & \frac{\alpha}{2} \|\mathbf{Z} - \mathbf{Z} \odot \mathbf{M}\|_F^2 + \langle \mathbf{Y}_1^k, \mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{E}^k \rangle + \langle \mathbf{Y}_2^k, \mathbf{Z} - \mathbf{J}^k \rangle \\ & + \langle \mathbf{Y}_3^k, \mathbf{Z} - \mathbf{L}^k \rangle + \frac{\mu^k}{2} (\|\mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{E}^k\|_F^2 + \|\mathbf{Z} - \mathbf{J}^k\|_F^2 + \|\mathbf{Z} - \mathbf{L}^k\|_F^2) \end{aligned} \tag{13}$$

Since the sub-problem for $\mathbf{Z}$ involves Hadamard product operator, which makes the problem hard to optimize. Alternatively, the $\mathbf{Z}$ in $\mathbf{Z} \odot \mathbf{M}$ can be obtained from former iteration. Thus, an iterative algorithm can be formed to solve the sub-problem of $\mathbf{Z}$, we have:

$$\begin{aligned}
\mathbf{Z}^{k+1} = \underset{\mathbf{Z}}{\operatorname{argmin}}\ & \frac{\alpha}{2}\|\mathbf{Z}-\mathbf{Z}^{k}\odot\mathbf{M}\|_F^2 + \langle \mathbf{Y}_1^k, \mathbf{X}-\mathbf{X}\mathbf{Z}-\mathbf{E}^k\rangle + \langle \mathbf{Y}_2^k, \mathbf{Z}-\mathbf{J}^k\rangle + \langle \mathbf{Y}_3^k, \mathbf{Z}-\mathbf{L}^k\rangle \\
& + \frac{\mu^k}{2}\left(\|\mathbf{X}-\mathbf{X}\mathbf{Z}-\mathbf{E}^k\|_F^2 + \|\mathbf{Z}-\mathbf{J}^k\|_F^2 + \|\mathbf{Z}-\mathbf{L}^k\|_F^2\right) \\
\mathbf{Z}^{k+1} = \underset{\mathrm{Z}}{\operatorname{argmin}}\ & \frac{\alpha}{2}\|\mathbf{Z}-\mathbf{Z}^{k}\odot\mathbf{M}\|_F^2 \\
& + \frac{\mu^k}{2}\left(\left\|\mathbf{X}-\mathbf{X}\mathbf{Z}-\mathbf{E}^k+\frac{\mathbf{Y}_1^k}{\mu^k}\right\|_{\mathrm{F}}^2 + \left\|\mathbf{Z}-\mathbf{J}^{\mathrm{k}}+\frac{\mathbf{Y}_2^k}{\mu^{\mathrm{k}}}\right\|_{\mathrm{F}}^2 + \left\|\mathbf{Z}-\mathbf{L}^{\mathrm{k}}+\frac{\mathbf{Y}_3^{\mathrm{k}}}{\mu^{\mathrm{k}}}\right\|_{\mathrm{F}}^2\right) \\
& = \left((\alpha/\mu^{\mathrm{k}}+2)\mathbf{I}+\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}(\mathbf{X}^{\mathrm{T}}(\mathbf{X}-\mathbf{E}^k)+\mathbf{J}^k+\mathbf{L}^k \\
& \quad + (\alpha(\mathbf{Z}^k\odot\mathbf{M})+\mathbf{X}^{\mathrm{T}}\mathbf{Y}_1^k-\mathbf{Y}_2^k-\mathbf{Y}_3^k)/\mu^k)
\end{aligned} \tag{14}$$

Step 2 (Update $\mathbf{E}$): $\mathbf{E}$ can be updated by solving following optimization problem (15):

$$\begin{aligned}
\mathbf{E}^{k+1} &= \underset{\mathbf{E}}{\operatorname{argmin}}\ \lambda\|\mathbf{E}\|_1 + \langle \mathbf{Y}_1^k, \mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}-\mathbf{E}\rangle + \frac{\mu^k}{2}\|\mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}-\mathbf{E}\|_F^2 \\
&= \underset{\mathbf{E}}{\operatorname{argmin}}\ \lambda\|\mathbf{E}\|_1 + \frac{\mu^k}{2}\left\|\mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}-\mathbf{E}+\frac{\mathbf{Y}_1^k}{\mu^k}\right\|_F^2 = S_{\lambda/\mu^k}\left(\mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}+\frac{\mathbf{Y}_1^k}{\mu^k}\right)
\end{aligned} \tag{15}$$

Step 3 (Update $\mathbf{J}$): $\mathbf{J}$ can be updated by solving the following optimization problem:

$$\mathbf{J}^{k+1} = \underset{\mathbf{J}}{\operatorname{argmin}}\ \|\mathbf{J}\|_* + \frac{\mu^k}{2}\left\|\mathbf{Z}^{k+1}-\mathbf{J}+\frac{\mathbf{Y}_2^k}{\mu^k}\right\|_F^2 = \mathbf{U}S_{1/\mu^k}[\mathbf{\Sigma}]\mathbf{V}^T \tag{16}$$

where $(\mathbf{U},\mathbf{\Sigma},\mathbf{V}^{\mathrm{T}}) = \mathrm{SVD}\left(\mathbf{Z}^{k+1}+\frac{\mathbf{Y}_2^k}{\mu^k}\right)$, $S_\varepsilon[\cdot]$ is the soft-thresholding (shrinkage) operator given by:

$$S_\varepsilon[\mathrm{x}] = \begin{cases} \mathrm{x}-\varepsilon, if\ \mathrm{x} > \varepsilon \\ \mathrm{x}+\varepsilon, if\ \mathrm{x} < -\varepsilon \\ 0, \text{otherwise} \end{cases} \tag{17}$$

Step 4 (Update $\mathbf{L}$): $\mathbf{L}$ can be updated by solving following optimization problem (18):

$$\mathbf{L}^{k+1} = \underset{\mathbf{L}}{\operatorname{argmin}}\ \beta\,\|\mathbf{L}\|_1 + \frac{\mu^k}{2}\left\|\mathbf{Z}^{k+1}-\mathbf{L}+\frac{\mathbf{Y}_3^k}{\mu^k}\right\|_F^2 = S_{\beta/\mu^k}\left(\mathbf{Z}^{k+1}+\frac{\mathbf{Y}_3^k}{\mu^k}\right) \tag{18}$$

Step 5: Update the multipliers and μ:

$$\begin{cases} \mathbf{Y}_1^{k+1} = \mathbf{Y}_1^k + \mu^k(\mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}-\mathbf{E}^{k+1}) \\ \mathbf{Y}_2^{k+1} = \mathbf{Y}_2^k + \mu^k(\mathbf{Z}^{k+1}-\mathbf{J}^{k+1}) \\ \mathbf{Y}_3^{k+1} = \mathbf{Y}_3^k + \mu^{\mathrm{k}}(\mathbf{Z}^{k+1}-\mathbf{L}^{k+1}) \\ \mu^{k+1} = \min(\rho\mu^k, \mu_{max}) \end{cases} \tag{19}$$

Algorithm 1. Solving BLSR by Inexact ALM

```
Input: Training data $\mathbf{X}$. Parameters $\alpha$, $\beta$ and $\lambda$.
Initialization: $\mathbf{Z}^0 = \mathbf{J}^0 = \mathbf{E}^0 = \mathbf{0}$, $\mathbf{Y}_1^0 = \mathbf{Y}_2^0 = \mathbf{Y}_3^0 = \mathbf{0}$,
    $\mu^0 = 10^{-5}$, $\mu_{max} = 10^8$, $\varepsilon = 10^{-6}$, $\rho = 1.1$.
While not converged do
Fix other variables and optimize $\mathbf{Z}^{k+1}$ via (14).
Fix other variables and optimize $\mathbf{E}^{k+1}$ via (15).
Fix other variables and optimize $\mathbf{J}^{k+1}$ via (16).
Fix other variables and optimize $\mathbf{L}^{k+1}$ via (18).
Update the multipliers and $\mu$ via (19).
Check the convergence conditions:
    $\|\mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}-\mathbf{E}^{k+1}\|_\infty < \varepsilon$, $\|\mathbf{Z}^{k+1}-\mathbf{L}^{k+1}\|_\infty < \varepsilon$
    $\|\mathbf{Z}^{k+1} - \mathbf{J}^{k+1}\|_\infty < \varepsilon$
8: **End while**
**Output**: **Z**
```

Generally, we outline the optimization process of BLSR in Algorithm 1. The major computational burden of BLSR is solving (14) and (16) because they involve matrix inversion and singular value decomposition (SVD). The overall computational complexity of BLSR is $\mathcal{O}(\tau(\mathrm{n}^2\mathrm{D} + \mathrm{n}^3))$, where τ is the iteration number, and n is the number of training samples.

3.2.2. Optimization for BLSGE

The trace-ratio problem in the form of (10) does not have a closed-form solution [27]. Consequently, such problem can be approximately solved as a determinant-ratio problem, so we turn to solve the following problem (20):

$$\min_{\mathbf{V}} \frac{|\mathbf{V}^{\mathrm{T}}\mathbf{X}\mathbf{L}^{\mathrm{intra}}\mathbf{X}^{\mathrm{T}}\mathbf{V}|}{|\mathbf{V}^{\mathrm{T}}\mathbf{X}\mathbf{L}^{\mathrm{inter}}\mathbf{X}^{\mathrm{T}}\mathbf{V}|} \tag{20}$$

With the method of Lagrangian multiplier, the solution of problem (20) is transformed to solve a generalized eigenvalues problem as follows:

$$\mathbf{X}\mathbf{L}^{\mathrm{intra}}\mathbf{X}^{\mathrm{T}}\mathbf{V} = \lambda\, \mathbf{X}\mathbf{L}^{\mathrm{inter}}\mathbf{X}^{\mathrm{T}}\mathbf{V} \tag{21}$$

Then we obtain the eigenvectors corresponding to the d minimum eigenvalues, and the projection matrix can be got as $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d]$. The detailed process of BLSGE is given in Algorithm 2. The complete procedure of BLSE is outlined in Algorithm 3.

```
**Algorithm 2. BLSGE**
**Input**: Affinity weights matrix **Z**, reduced dimension $d$.
1: Compute the weights of inter-class graph (6) and intra-class graph (7) through affinity matrix **Z**.
2: Solve the generalized eigenvalue problem (21), and get the eigenvectors corresponding to the $d$ minimum eigenvalues.
**Output**: Projection matrix **V**.
```

```
**Algorithm 3. BLSE**
**Input**: labeled training data $\mathbf{X} \in \Re^{\mathrm{D} \times n}$. Reduced dimension $d$.
        Tradeoff parameters $\alpha$, $\beta$ and $\gamma$.
1: Run **Algorithm 1** to get the affinity weights matrix **Z** of **X**.
2: Run **Algorithm 2** to obtain the optimal projection matrix **V**.
**Output**: Projection matrix **V**.
```

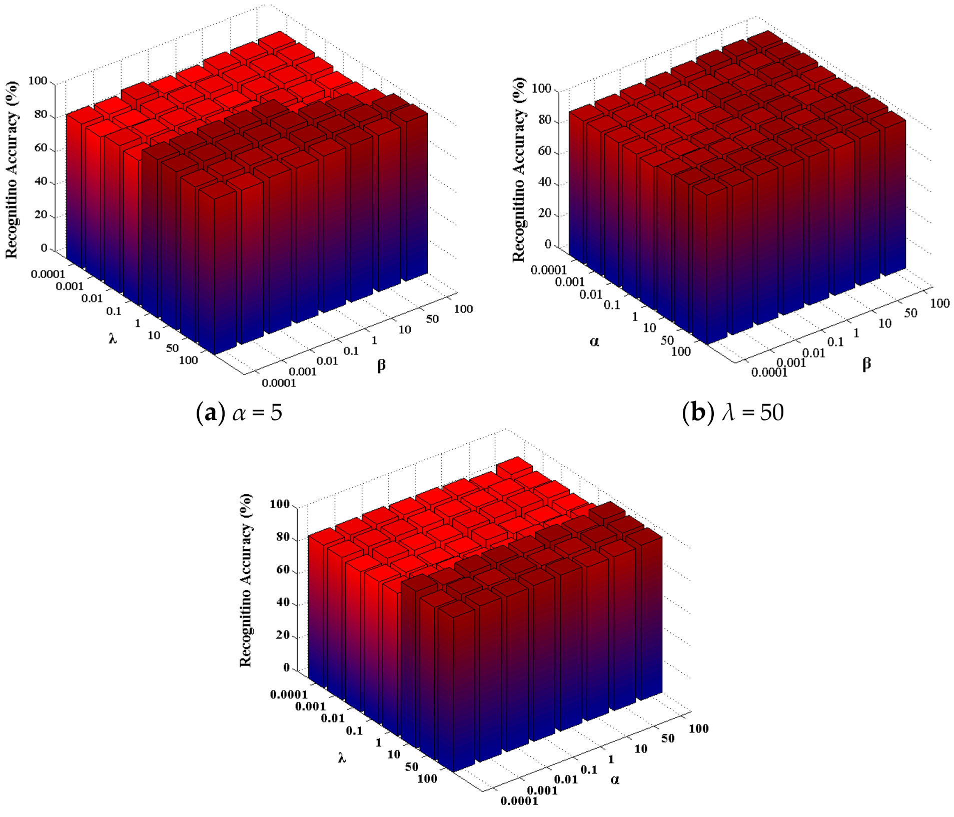
3.3. Classification

For classification, we directly use the calculated projection **V** to obtain the transformation results of the training and testing data. One can apply existing classifier such as 1-Nearest Neighbor (NN) to classify the projected results of testing data.

4. Experimental Results

4.1. Analysis of BLSE

The representation results of BLSR have a great influence on the graph construction and the performance of BLSE. There are three parameters in BLSR, i.e., regularization parameters α, β and λ. λ and β controls the sparsity noise term **E** and representation **Z**. α is used to regularize the representation to be block-diagonal. We conduct experiments to study the sensitivity of the proposed BLSE over a wide range of these parameters. ORL data [42] are divided into training and testing samples for tuning these parameters, and the reduced dimension is 50. The experimental results obtained are used to find an effective range of parameters to ensure a reliable performance. The results are reported in Figure 4. From the results, one can observe that the performance of BLSE is not especially sensitive to λ and is robust in a quiet large range for α and β.



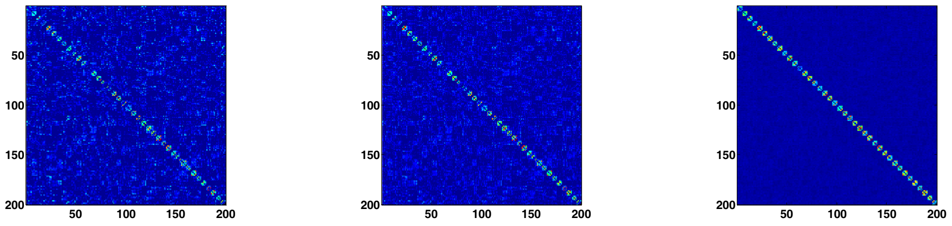
(**a**) $\alpha = 5$

(**b**) $\lambda = 50$

(**c**) $\beta = 50$

Figure 4. Sensitivity analysis of α, β and λ on ORL dataset. (**a**) Tuning β and λ with α fixed; (**b**) tuning α and β with λ fixed; (**c**) tuning α and λ with β fixed.

Figure 5 further shows the graph weights matrix obtained by BLSR and the corresponding recognition accuracy obtained by BLSE with $\alpha = 1$, 10, and 100 respectively. A larger penalty will be imposed on the block-diagonal regularization term as α increases. The obtained graph weights matrix tends to show a better block-diagonal structure. Benefitting from the block-diagonal graph weights matrix, BLSE achieves higher recognition accuracy. The results demonstrate that the proposed method can enhance the block-diagonal structure of graph weights matrix, which helps achieve better recognition performance. However, an extremely large α will regularize the inter-class representation in $\mathbf{Z}$ to be zero, which might not be able to reveal the inter-class adjacent relationship among samples well. To achieve reliable and stable performance, a suggested parameter settings are 100 > α >1, 50 > β > 0.01, 50 > λ >1.

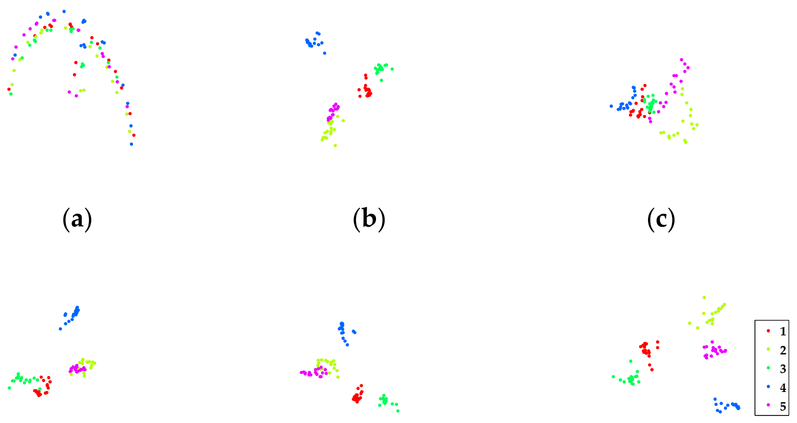


(**a**) $\alpha = 1$ (Acc = 93.85%) (**b**) $\alpha = 10$ (Acc = 94.18%) (**c**) $\alpha = 100$ (Acc = 94.35%)

Figure 5. Visualization of the graph weights and corresponding recognition accuracy obtained by BLSE. From left to right, α is 1, 10 and 100, respectively.

4.2. 2-D Visualization Experiment on CMU PIE Dataset

In this part, a partial CMU PIE face database [43] (120 images of five persons) is used to intuitively show the discriminate ability of different methods using t-SNE [44]. In the experiment, seven images per person are randomly selected for training, and the remaining about 17 images for testing. Figure 6a–g visualize the testing data distributions along the first two dimensions acquired by different methods. From Figure 6, we may draw several conclusions. First, as classical supervised DR methods, LDA and MMC can yield superior performance to that of unsupervised PCA. Second, SGDA [27] only exploits the local neighborhood structure via sparse representation, and it does not perform well, as shown in Figure 6d. Some parts of class 1, 2, 3 and 4 mix together. By introducing global low-rank regularization, LGDA [28] shows better separation ability. Nevertheless, there are overlaps between class 2 and class 5, class 1 and class 3. With both sparse and low-rank constraints, SLGDA [28] performs better than SGDA and LGDA. However, class 2 and class 5 still have significant overlaps as shown in Figure 6f. Contrastively, the proposed BLSE successfully separates all the classes with clear boundaries between them, which can be explained by the simultaneously imposed local sparse, global low-rank and the discriminative block-diagonal structure constraint. The experiment shows that BLSE has the capacity to separate complex face data distribution.



(**a**) (**b**) (**c**) (**d**)

(**e**) (**f**) (**g**)

Figure 6. Two-dimensional five-class CMU PIE data projected by different DR methods. (**a**) PCA; (**b**) LDA; (**c**) MMC; (**d**) SGDA; (**e**) LGDA; (**f**) SLGDA; (**g**) BLSE.

4.3. Experimental Results on Image Datasets

We conducted extensive experiments to evaluate the performance of the proposed method on widely used face and object databases (ORL [42], Yale [45], CMU PIE [43], and COIL 20 [46]). Figure 7 visually demonstrates characteristics of each database. Our approach is compared with several state-of-the-art subspace learning approaches including PCA, LDA, MMC [6], SGDA [27], LGDA [28], and SLGDA [28]. To make the comparison fair, for all the evaluated algorithms we first apply PCA as preprocessing step by retaining 99% energy. A nearest neighbor classifier is employed in the projected feature space for all the methods.

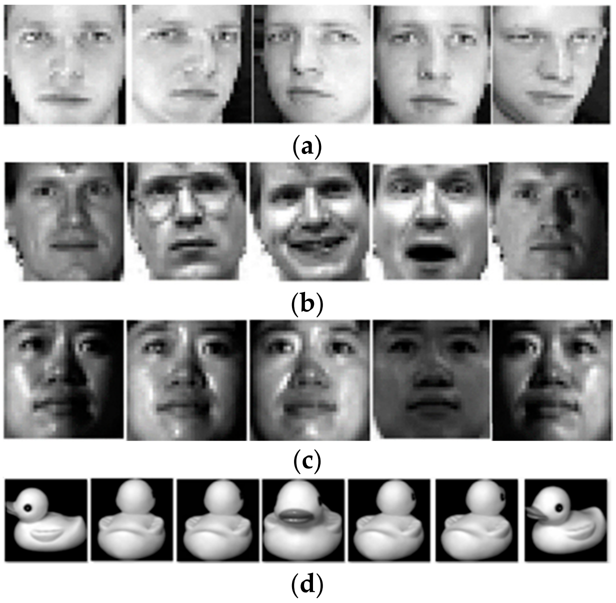

(**a**)
(**b**)
(**c**)
(**d**)

Figure 7. The sample images of public datasets used in our experiment. (**a**) The ORL dataset; (**b**) The Yale database; (**c**) The CMU PIE dataset; (**d**) The COIL20 dataset.

The ORL face database consists of a total of 400 face images from 40 individuals with 10 images per person. The images were taken at different times, lighting variation, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glassed/no glassed) against a dark homogeneous background. In the experiments, each image in ORL database is manually cropped and resized to 32 × 32. Using the five samples per person from ORL database as training set, we present the first five basis vectors of Eigenfaces, Fisherfaces, and our BLSEFaces in Figure 8.

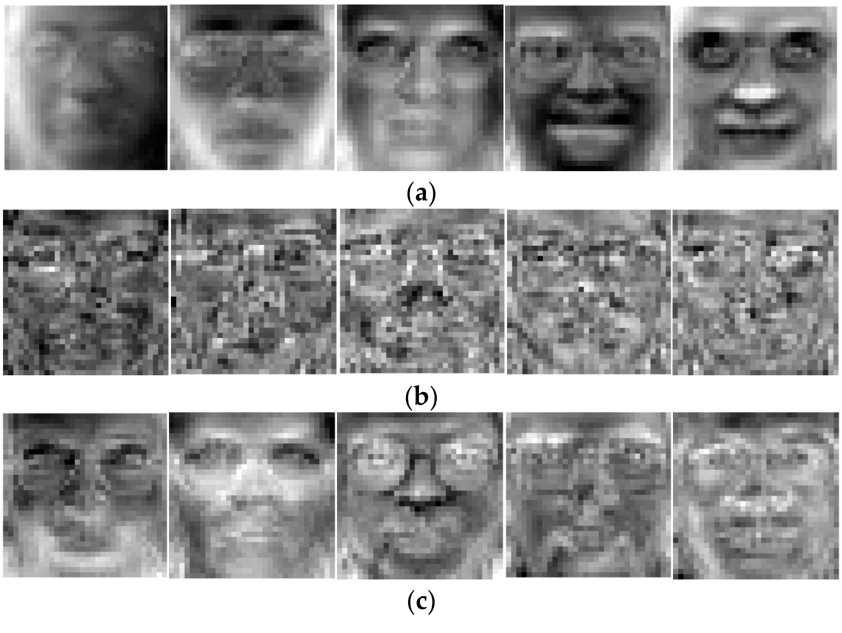

(**a**)
(**b**)
(**c**)

Figure 8. The first five basis vectors calculated by (**a**) PCA (Eigenfaces); (**b**) LDA (Fisherfaces); (**c**) BLSE (BLSEFaces) on ORL dataset.

A random subset with *t* (=3, 4, 5, 6) images of each individual is selected for training and the rest for testing. For each *t*, we run the programs 10 times and calculate the recognition rates as well as the standard deviations with different reduced dimensions.

The Yale face database contains 165 gray scale images of 15 individuals, each individual has 11 images. The images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink). In our experiments, each image in Yale database was manually cropped and resized to 32 × 32. A random subset with *t* (=4, 5, 6, 7) images each individual is selected for learning the embedding and the rest for testing. For each giving *t*, we run each program 10 times to randomly choose the training set and report the average recognition rates as well as the standard deviations with different reduced dimensions.

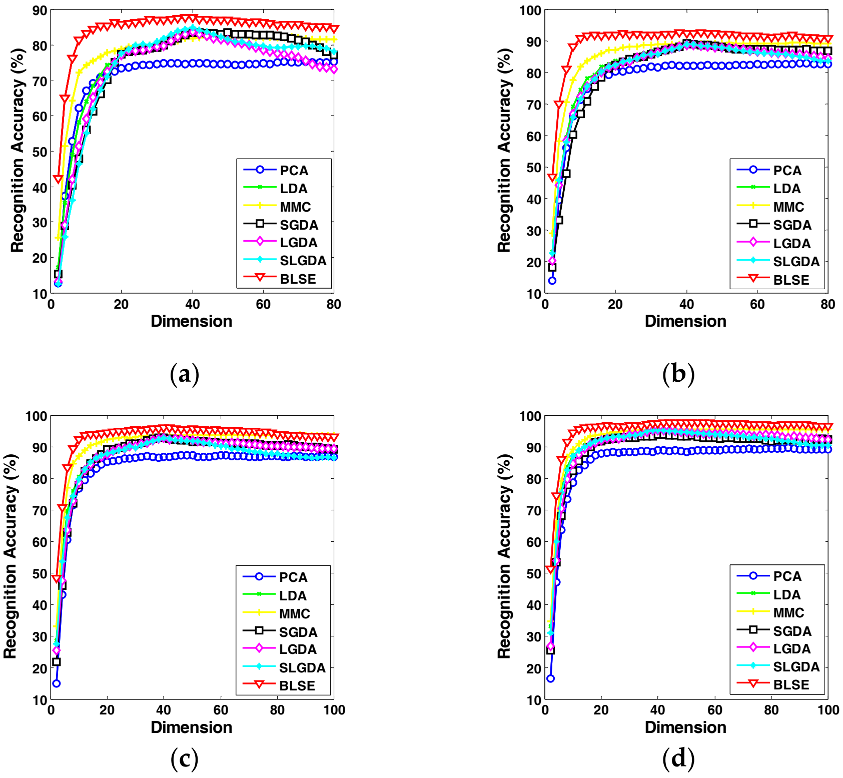
The CMU PIE dataset contains over 40,000 face images of 68 individuals. Images of each individual were acquired across 13 different poses under 43 different illumination conditions, and with four different expressions. Here we use a near frontal pose subset, namely C07, for experiments, which contains 1629 images of 68 individuals. Each individual has about 24 images. All images are manually cropped and resized to 32 × 32 pixel. A random subset with *t* (=4, 5, 6, 7) images for each individual is selected for learning the embedding and the rest for testing. For each giving *t*, we perform 10 times to randomly choose the training set and report the average recognition rates as well as the standard deviations under different dimensions.

The COIL20 image dataset contains 1440 gray scale images of 20 objects with 72 images per subject. The images of each object were taken 50 apart as the object was rotated on a turntable. Each image is of size 32 × 32. Following the experimental setting in [47], we selected the first 36 images per subject for training and the remaining images for testing in this experiment.

Figures 9–12 plot the curves of average recognition accuracy versus different dimensions on ORL, Yale, CMU PIE and COIL 20 databases, respectively. Moreover, the details of experiments results, namely maximal recognition rates together with the standard deviation and dimension of different algorithms are summarized in Table 1.

Table 1. Recognition rates (%) and the corresponding standard deviations and dimensions (in parenthesis) on ORL, Yale, CMU PIE and COIL20 databases.

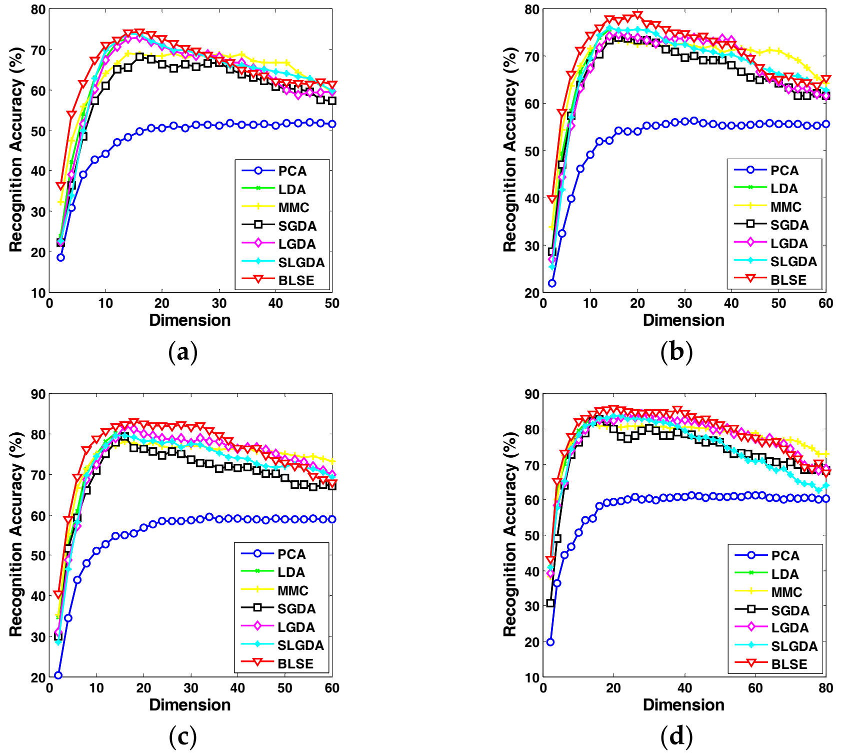
Dataset	*t*	Compared Methods						Ours
		PCA	LDA	MMC	SGDA	LGDA	SLGDA	BLSE
ORL	3	75.32 ± 2.58 (80)	83.63 ± 2.38 (38)	82.39 ± 2.73 (56)	83.61 ± 2.72 (40)	83.79 ± 2.15 (40)	84.96 ± 1.78 (40)	**87.68 ± 2.36** **(38)**
	4	82.92 ± 1.79 (72)	88.71 ± 2.26 (38)	89.58 ± 2.44 (72)	89.42 ± 2.08 (40)	88.83 ± 2.36 (40)	89.08 ± 2.63 (40)	**92.63 ± 2.28** **(40)**
	5	87.45 ± 1.70 (48)	93.15 ± 1.13 (38)	93.95 ± 1.55 (64)	93.10 ± 1.94 (40)	93.00 ± 0.91 (40)	92.95 ± 1.41 (40)	**96.05 ± 0.96** **(42)**
	6	89.75 ± 1.65 (86)	95.75 ± 1.31 (38)	95.81 ± 1.22 (70)	94.06 ± 1.48 (40)	95.50 ± 1.09 (40)	95.44 ± 1.22 (40)	**97.88 ± 0.67** **(44)**
Yale	4	52.00 ± 2.34 (46)	72.76 ± 2.30 (14)	69.05 ± 3.51 (14)	68.29 ± 3.81 (16)	72.86 ± 2.16 (16)	73.90 ± 2.92 (14)	**74.38 ± 2.48** **(16)**
	5	56.33 ± 5.57 (32)	75.67 ± 2.31 (14)	73.44 ± 4.52 (14)	74.00 ± 3.52 (16)	74.56 ± 3.33 (16)	76.11 ± 2.83 (14)	**78.89 ± 3.10** **(20)**
	6	59.60 ± 5.76 (34)	80.27 ± 3.81 (14)	78.13 ± 5.56 (16)	79.47 ± 3.51 (16)	81.87 ± 3.51 (16)	79.60 ± 5.34 (14)	**83.20 ± 4.71** **(18)**
	7	61.33 ± 5.02 (42)	83.00 ± 2.70 (14)	81.67 ± 4.30 (14)	82.83 ± 3.34 (16)	84.17 ± 4.10 (28)	83.83 ± 4.45 (22)	**85.83 ± 3.17** **(20)**
CMU PIE	4	53.72 ± 1.28 (100)	91.14 ± 1.25 (64)	88.61 ± 1.50 (96)	89.44 ± 1.35 (100)	91.33 ± 0.81 (86)	90.40 ± 1.03 (94)	**92.28 ± 1.07** **(80)**
	5	60.31 ± 1.78 (100)	92.60 ± 0.83 (66)	91.27 ± 0.85 (98)	89.95 ± 1.84 (98)	92.60 ± 0.83 (66)	91.82 ± 0.99 (100)	**93.26 ± 1.03** **(92)**
	6	65.88 ± 1.92 (120)	93.56 ± 0.88 (66)	93.17 ± 1.03 (106)	92.15 ± 0.95 (112)	93.24 ± 1.04 (104)	92.84 ± 0.92 (118)	**93.93 ± 1.01** **(92)**
	7	71.12 ± 1.61 (120)	94.34 ± 0.83 (66)	94.09 ± 0.66 (110)	93.54 ± 0.58 (120)	94.09 ± 0.72 (104)	93.88 ± 0.59 (102)	**94.64 ± 0.59** **(70)**
COIL20	36	86.39 (28)	88.75 (14)	90.97 (12)	78.33 (26)	87.78 (12)	90.00 (90)	**92.22 (8)**



(**a**) (**b**)

(**c**) (**d**)

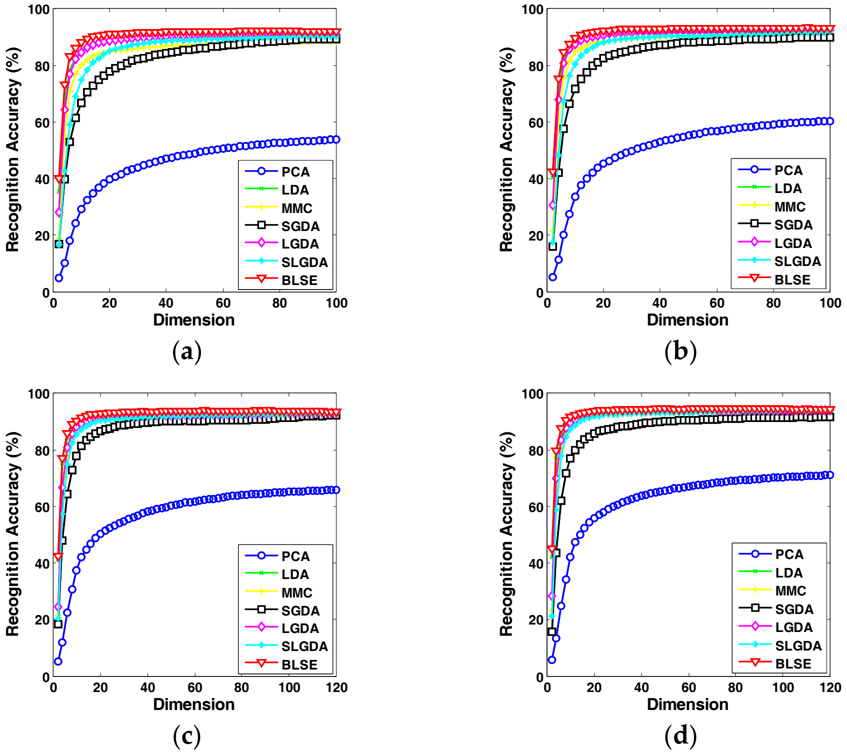
Figure 9. Recognition rate (%) versus dimension of different methods on the ORL database. (**a**) Three training samples per person; (**b**) Four training samples per person; (**c**) Five training samples per person; (**d**) Six training samples per person.



(**a**) (**b**)

(**c**) (**d**)

Figure 10. Recognition rate (%) versus dimension of different methods on the Yale dataset. (**a**) Four training samples per person; (**b**) Five training samples per person; (**c**) Six training samples per person; (**d**) Seven training samples per person.



(**a**) (**b**) (**c**) (**d**)

Figure 11. Recognition rate (%) versus dimension of different methods on the CMU PIE dataset. (**a**) Four training samples per person; (**b**) Five training samples per person; (**c**) Six training samples per person; (**d**) Seven training samples per person.

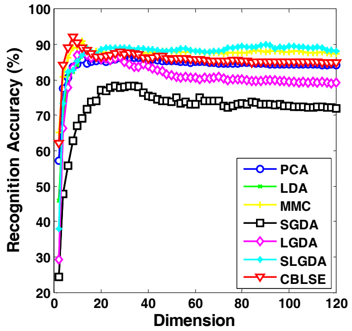


Figure 12. Recognition rate (%) versus dimension of different methods on the COIL20 database.

5. Discussion and Conclusions

Based on the experimental results on face and object image datasets, one can conclude that with the increase of dimensions and the number of training samples per class, all the methods tend to achieve better performance. PCA is simple to calculate, and performs well in some cases, but its unsupervised nature restricts its performance. By introducing supervised information with different discrimination criteria, LDA and MMC can achieve better performance. SGDA, LGDA and SLGDA can adaptively select neighbors for graph construction, and find the representation of each sample using the labeled samples in the same class to purse block-diagonal structure representations. However, this process may result in large representation error due to the limited samples per class, which might not be able to reveal the intra-class adjacent relationship well. Besides, SGDA, LGDA and SLGDA disconnect inter-class samples in graph construction. The operation cannot capture the inter-class adjacent relationship well. As a result, SGDA, LGDA and SLGDA do not perform well, as the experimental results show. Comparably, the proposed BLSE model can achieve better performance. The reason is twofold. Firstly, the developed BLSR method can capture both local neighborhood relations and global structures latent in data with low-rank and sparse constraints. Different from SGDA, LGDA and SLGDA, all samples are employed in BLSR when finding the representation of each sample. The introduction of block-diagonal regularization can capture the intra-class and inter-class adjacent relationships hidden in data, and enhance the identification capability of BLSR. Secondly, benefit from BLSR and GE framework, the discriminative capacity of low dimensional subspace learned by BLSGE is further boosted by simultaneously minimizing the intra-class scatter and maximizing the inter-class scatter.

To conclude, we have proposed a novel block-diagonal constrained low-rank and sparse based embedding (BLSE) model for the dimensionality reduction and classification of image data. Two procedures of BLSE, namely, block-diagonal constrained low-rank and sparse representation (BLSR) and block-diagonal constrained low-rank and sparse graph embedding (BLSGE), are detailed. BLSR takes the advantages of local discriminative capacity of SR and the global low-rank property of LRR. Meanwhile, a novel block-diagonal regularization term is introduced to fully harness the label information and purse a block-diagonal representation. The affinity weights matrix obtained by BLSR can well reveal the intra-class similarities and inter-class differences of data. With the intra-class and inter-class graphs derived from BLSR, BLSGE finds a low-dimensional subspace with enhanced intra-class compactness and inter-class separation. Experimental results on public face and object datasets are performed, and validate the effectiveness of BLSE model.

Acknowledgments: The authors would like to thank the anonymous reviewers and the editors for their valuable suggestions. This work was supported by the National Natural Science Foundation of China (No. 61571069, No. 61401048), Chongqing University Postgraduates' Innovation Project (No. CYB15030), and in part by the Fundamental Research Funds for the Central Universities (No. 106112017CDJQJ168817, No. 106112017CDJQJ168819).

Author Contributions: Tan Guo, Xiaoheng Tan and Lei Zhang conceived and designed the global structure and methodology of the dissertation; Chaochen Xie and Lu Deng provided some valuable advice and proofread the manuscript. Tan Guo analyzed the data and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Reference

1. Jain, A.; Duin, R.; Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 4–37.
2. Wang, X.; Zheng, Y.; Zhao, Z.; Wang, J. Bearing fault diagnosis based on statistical locally linear embedding. *Sensors* **2015**, *15*, 16225–16247.
3. Zhao, X.; Zhang, S. Facial expression recognition based on local binary patterns and kernel discriminant Isomap. *Sensors* **2011**, *11*, 9573–9588.
4. Jolliffe, I.T. *Principal Component Analysis*; Springer: New York, NY, USA, 1986.
5. Webb, A.R.; Copsey, K.D. *Introduction to Statistical Pattern Recognition*, 3nd ed.; John Wiley & Sons. Ltd.: Hoboken, NJ, USA, 1990.
6. Li, H.; Jiang, T.; Zhang, K. Efficient and robust feature extraction by maximum margin criterion. *IEEE Trans. Neural Netw.* **2006**, *17*, 1157–1165.
7. Schölkopf, B.; Smola, A.; Müller, K.R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1319.
8. Mika, S.; Ratsch, G.; Weston, J.; Scholkopf, B.; Smola, A.; Muller, K.R. Constructing descriptive and discriminative nonlinear features: Rayleigh coefficients in kernel feature spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 623–628.
9. Roweis, S.T.; Saul, L.K. Nonlinear dimension reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326.
10. Tenenbaum, J.B.; De Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323.
11. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396.

12. Yan, S.; Xu, D.; Zhang, B.; Zhang, H.J.; Yang, Q.; Lin, S. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 40–51.
13. Han, Y.; Jin, A.T.B.; Abas, F.S. Neighborhood preserving discriminant embedding in face recognition. *J. Visual Commun. Image Represent.* **2009**, *20*, 532–542.
14. Huang, H.; Luo, F.; Liu, J.; Yang, Y. Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding. *ISPRS J. Photogramm. Remote Sens.* **2015**, *106*, 42–54.
15. Qiao, L.; Chen, S.; Tan, X. Sparsity preserving projections with applications to face recognition. *Pattern Recognit.* **2010**, *43*, 331–341.
16. Elhamifar, E.; Vidal, R. Sparse subspace clustering: algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2765–2781.
17. Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; Ma, Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 171–184.
18. Cheng, B.; Yang, J.; Yan, S.; Fu, Y.; Huang, T.S. Learning with l_1-graph for image analysis. *IEEE Trans. Image Process.* **2010**, *19*, 858–866.
19. Liu, G.; Liu, Q.; Li, P. Blessing of dimensionality: recovering mixture data via dictionary pursuit. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 47–60.
20. Lu, C.Y.; Min, H.; Zhao, Z.Q.; Zhu, L.; Huang, D.S.; Yan, S. Robust and efficient subspace segmentation via least squares regression. In Proceedings of the European Conference on Computer Vision (ECCV), Firenze, Italy, 7–13 October 2012.
21. Zhao, H.; Ding, Z.; Fu, Y. Block-wise constrained sparse graph for face image representation. In Proceedings of the International Conference on Automatic Face and Gesture Recognition, Ljubljana, Slovenia, 4–8 May 2015.
22. Zhao, H.; Ding, Z.; Fu, Y. Ensemble subspace segmentation under sparse and block-wise constraints. *IEEE Trans. Circuits Syst. Video Tech.* **2017**, doi:10.1109/TCSVT.2017.2678443.
23. Tang, K.; Liu, R.; Su, Z.; Zhang, J. Structure-constrained low-rank representation. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 2167–2179.
24. Zhang, Y.; Jiang, Z.; Davis, L.S. Learning structured low-rank representations for image classification. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR), Portland, OR, USA, 23–28 June 2013.
25. Li, Y.; Liu, J.; Lu, H.; Ma, S. Learning robust face representation with classwise block-diagonal structure. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2051–2062.
26. Feng, J.; Lin, Z.; Xu, H.; Yan, S. Robust subspace segmentation with block-diagonal prior. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR), Columbus, OH, USA, 23–28 June 2014.
27. Ly, N.H.; Du, Q.; Fowler, J.E. Sparse graph-based discriminant analysis for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3872–3884.
28. Li, W.; Liu, J.; Du, Q. Sparse and low-rank graph for discriminant analysis of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4094–4105.
29. Zhuang, L.; Gao, H.; Lin, Z.; Ma, Y.; Zhang, X.; Yu, N. Non-negative low rank and sparse graph for semi-supervised learning, In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (IEEE CVPR), Providence, RI, USA, 16–21 June 2012.
30. Zhao, M.; Jiao, L.; Feng, J.; Liu, T. A simplified low rank and sparse graph for semi-supervised learning. *Neurocomputing* **2014**, *140*, 84–96.
31. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227.
32. Cai, J.; Chen, J.; Liang, X. Single-sample face recognition based on intra-class differences in a variation model. *Sensors* **2015**, *15*, 1071–1087.
33. Zhao, H.; Ding, Z.; Fu, Y. Pose-dependent low-rank embedding for head pose estimation. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
34. Zhang, L.; Zhang, D. Robust visual knowledge transfer via extreme learning machine based domain adaptation. *IEEE Trans. Image Process.* **2016**, *25*, 4959–4973.
35. Zhang, L.; Zhang, D. Evolutionary cost-sensitive extreme learning machine. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, doi:10.1109/TNNLS.2016.2607757.

36. Zhang, L.; Zhang, D. Visual understanding via multi-feature shared learning with global consistency. *IEEE Trans. Multimed.* **2016**, 18, 247–259.
37. Zhang, Z.; Li, Y.; Wang, F.; Meng, G.; Salman, W.; Saleem, L. A novel multi-sensor environmental perception method using low-rank representation and a particle filter for vehicle reversing safety. *Sensors* **2016**, *16*, 848.
38. Xu, Y.; Fang, X.; Wu, J.; Li, X.; Zhang, D. Discriminative transfer subspace learning via low-rank and sparse representation. *IEEE Trans. Image Process.* **2015**, *25*, 850–863.
39. Zhang, L.; Zuo, W.; Zhang, D. LSDT: Latent sparse domain transfer learning for visual adaptation. *IEEE Trans. Image Process*. **2016**, *25*, 1179–1191.
40. Guo, T.; Zhang, L.; Tan, X. Neuron pruning-based discriminative extreme learning machine for pattern classification. *Cogn. Comput*. **2017**, doi: 10.1007/s12559-017-9474-4.
41. Lin, Z.; Liu, R.; Su, Z. Linearized alternating direction method with adaptive penalty for low rank representation, In Proceedings of the 2011 Advances in Neural Information Processing Systems (NIPS), Granada, Spain, 12–17 December 2011.
42. He, X.; Yan, S.; Hu, Y.; Niyogi, P.; Zhang, H. Face recognition using Laplacian faces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 328–340.
43. Sim, T.; Baker, S.; Bsat, M. The CMU pose, illumination and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1615–1618.
44. Laurens, V.D.M.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
45. Cai, D.; He, X.; Han, J.; Zhang, H.-J. Orthogonal Laplacian faces for face recognition. *IEEE Trans. Image Process.* **2006**, *15*, 3608–3614.
46. Cai, D.; He, X.; Han, J.; Huang, T. Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2011, 33, 1548–1560.
47. Miao, S.; Wang, J.; Gao, Q.; Chen, F.; Wang, Y. Discriminant structure embedding for image recognition. *Neurocomputing* **2016**, *174*, 850–857.