

Research Article

Subspace Clustering with Sparsity and Grouping Effect

Binbin Zhang, Weiwei Wang, and Xiangchu Feng

School of Mathematics and Statistics, Xidian University, Xi'an 710126, China

Correspondence should be addressed to Weiwei Wang; wwwang@mail.xidian.edu.cn

Received 6 December 2016; Accepted 6 March 2017; Published 22 March 2017

Academic Editor: Simone Bianco

Copyright © 2017 Binbin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Subspace clustering aims to group a set of data from a union of subspaces into the subspace from which it was drawn. It has become a popular method for recovering the low-dimensional structure underlying high-dimensional dataset. The state-of-the-art methods construct an affinity matrix based on the self-representation of the dataset and then use a spectral clustering method to obtain the final clustering result. These methods show that sparsity and grouping effect of the affinity matrix are important in recovering the low-dimensional structure. In this work, we propose a weighted sparse penalty and a weighted grouping effect penalty in modeling the self-representation of data points. The experimental results on Extended Yale B, USPS, and Berkeley 500 image segmentation datasets show that the proposed model is more effective than state-of-the-art methods in revealing the subspace structure underlying high-dimensional dataset.

1. Introduction

Many high-dimensional datasets such as face images, motion trajectories, and text can be well approximated by multiple low-dimensional subspaces [1]. Recovering the low-dimensional structure underlying these datasets leads to the subspace clustering problem. It aims to group a set of data from a union of subspaces into the subspace (or cluster) from which it was drawn. It has shown promising applications in face clustering [2], motion segmentation [3], image segmentation [4], and system identification [5].

Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in R^{d \times n}$ be a set of sufficient sampled data vectors, where \mathbf{X} is the data matrix with each column corresponding to a data point, d is the feature dimension, and n is the number of data vectors. Assume that the data points are drawn from a union of k subspaces $\{S_i\}_{i=1}^k$ of unknown dimensions, respectively. Subspace segmentation aims to cluster the data into the underlying subspace from which they are drawn. Figure 1 presents an example of the subspace clustering problem.

1.1. Related Works. Most recent works focus on the spectral clustering-based methods [6–12]. They first learn an affinity matrix capturing the similarity between pairs of sample points from the dataset and then use a spectral clustering method such as the normalized cut (Ncut) [13] to obtain

the final clustering result. The affinity matrix is generally learnt by using the self-representation of the dataset under the assumption that each data point can be approximated by a linear or affine combination of other data points. To ensure that the affinity matrix has some expected properties, the previous methods [6–12] solve the following minimization problem:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \quad & \|\mathbf{Z}\|_p + \lambda \|\mathbf{E}\|_q \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \\ & \text{diag}(\mathbf{Z}) = \mathbf{0}, \end{aligned} \quad (1)$$

where $\|\mathbf{Z}\|_p$ is a proper norm of the representation matrix $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n] = [Z_{ij}] \in R^{n \times n}$ and is used to enforce \mathbf{Z} to have expected property, $\|\mathbf{E}\|_q$ measures the error matrix \mathbf{E} , noise, or corruption, and λ is a tradeoff parameter. The constraint $\text{diag}(\mathbf{Z}) = \mathbf{0}$ is optional and it is used to avoid the trivial solution. Once the optimal solution \mathbf{Z}^* of (1) is obtained, it is used to construct the affinity matrix \mathbf{A} . By the theory of spectral clustering, the affinity matrix which measures the similarity between data points should be nonnegative and symmetrical. A commonly used definition is

$$\mathbf{A} = \frac{(|\mathbf{Z}^*| + |\mathbf{Z}^{*T}|)}{2}, \quad (2)$$

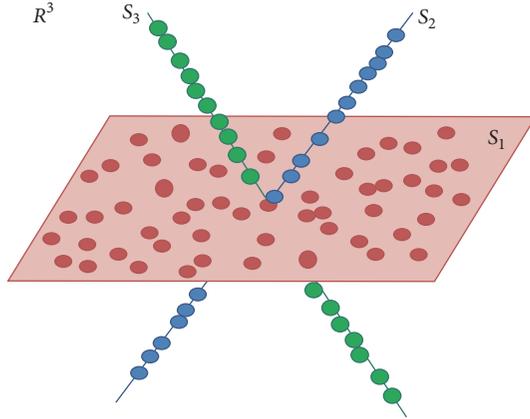


FIGURE 1: An example of subspace clustering.

where $|\cdot|$ denotes the element-wise absolute value of the matrix \mathbf{Z}^* . The method of sparse subspace clustering (SSC) [6, 7] pursues the sparsest representation for each data point by using the l_1 -norm regularization. Low-rank representation (LRR) [2, 8] seeks for the lowest-rank representation of all data points by employing the nuclear norm $\|\cdot\|$ of the coefficient matrix, which can capture the global structures of the data. Least square regression (LSR) [9] uses the Frobenius norm regularization. The correlation adaptive subspace segmentation (CASS) [10] uses the trace Lasso defined by $\|\mathbf{X} \text{diag}(\mathbf{z})\|_*$ to regularize each column of \mathbf{Z} to gain a tradeoff effect between l_1 -norm and l_2 -norm depending on the correlation of data points. In particular, when the data are highly correlated (i.e., $\mathbf{X}^T \mathbf{X}$ is close to $\mathbf{1}\mathbf{1}^T$), it is close to the l_2 -norm, while when the data are almost uncorrelated (i.e., $\mathbf{X}^T \mathbf{X}$ is close to \mathbf{I}), it behaves like the l_1 -norm. The smooth representation (SMR) [11] and SSC with the weighting function (W-SSC) [12] use the spatial information to design the regularization term of \mathbf{Z} to improve the clustering results.

The previous works pursue two properties of the self-representation and the resulting affinity matrix to ensure the success of subspace clustering. SSC [6, 7] and W-SSC [12] pursue sparsity. LRR [2], LSR [9], CASS [10], and SMR [11] are shown to have the grouping effect (GE, Definition 1), which tend to encourage clustering highly correlated data together. Under some ideal conditions, sparsity results in zero connection between data points from different subspaces and tends to group them into different clusters. However, SSC has the instability problem: if the data from the same subspace are highly correlated or clustered, it will only select one of the several related pieces of data at random and ignore other correlated data. Moreover, the datasets in practice not necessarily satisfy the ideal conditions. These limit the performance of SSC. W-SSC compresses the coefficients corresponding to data points spatially far away from the data in consideration by introducing a weight depending on the spatial distance between data points. However, SSC and W-SSC do not take the GE into consideration and both have the instability problem. LRR [2], LSR [9], CASS [10], and SMR [11] have the GE but they are insufficient in enforcing the sparsity.

1.2. Contributions. In this work, we pursue both sparsity and grouping effect in modeling the self-representation of data points. Specifically, for the representation vector \mathbf{z}_j of a data point \mathbf{x}_j , the sparsity is enforced by minimizing the l_1 -norm of coefficients only corresponding to data which is far away from \mathbf{x}_j . We also seek for GE by enforcing close representation vector for close data.

We present an algorithm based on the Alternating Direction Method of Multipliers (ADMM) to solve our model and two subproblems have analytical solution.

We test our method on the face dataset Extended Yale B and handwritten digit dataset USPS. The experimental results show that the proposed model is more effective than state-of-the-art methods in revealing the cluster structure underlying these high-dimensional datasets. We also test our method on Berkeley 500 image segmentation dataset for image segmentation; both the visual segmentation effects and segmentation metrics show that our method is better than state-of-the-art methods.

2. Self-Representation with Sparsity and Grouping Effect

2.1. Modeling. Let G_j be the set of indices i such that \mathbf{x}_i and \mathbf{x}_j are drawn from the same subspace or cluster and consider the representation $\mathbf{z}_j = (Z_{1j} \ Z_{2j} \ \cdots \ Z_{nj})^T$ ($Z_{jj} = 0$) (the j th column of \mathbf{Z}) for the data \mathbf{x}_j :

$$\mathbf{x}_j = \mathbf{X}\mathbf{z}_j = \sum_i Z_{ij} \mathbf{x}_i = \sum_{i \in G_j} Z_{ij} \mathbf{x}_i + \sum_{i \notin G_j} Z_{ij} \mathbf{x}_i \quad (3)$$

To make zero connection between data points from different subspaces, only $\sum_{i \notin G_j} |Z_{ij}|$ should be minimized; here $|Z_{ij}|$ denotes the absolute value of Z_{ij} . However, in the scenario of subspace clustering, G_j is unknown in advance. SSC pursues sparsity by minimizing $\sum_{i \neq j} |Z_{ij}| = \sum_{i \in G_j} |Z_{ij}| + \sum_{i \notin G_j} |Z_{ij}|$. So it is possible for SSC to make $Z_{ij} = 0$ if $i \in G_j$ and cause the instability problem. It is also possible that $Z_{ij} \neq 0$ if $i \notin G_j$, thus causing a nonzero connection between clusters. W-SSC compresses the coefficient Z_{ij} corresponding to the data points \mathbf{x}_i by introducing a weight depending on the spatial distance between \mathbf{x}_i and \mathbf{x}_j . However, SSC and W-SSC do not take the GE into consideration and both have the instability problem.

In [11], the authors define the following ‘‘grouping effect.’’ The works in [9–11] confirm that the grouping effect tends to group highly correlated data into the same cluster and experiments show that the grouping effect improves the clustering results on several commonly used datasets [14, 15].

Definition 1 (grouping effect [11]). Given a set of d -dimensional data points $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in R^{d \times n}$, a subspace representation matrix $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n] \in R^{n \times n}$ has grouping effect if $\|\mathbf{z}_i - \mathbf{z}_j\|_2 \rightarrow 0$ whenever $\|\mathbf{x}_i - \mathbf{x}_j\|_2 \rightarrow 0$, $\forall i \neq j$.

Taking both the sparsity and GE into consideration of the self-representation of a data point, we propose to solve the following minimization problem:

$$\begin{aligned} \min_{z_j} \quad & \sum_i S_{ij} |Z_{ij}| + \frac{\lambda}{2} \sum_i W_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 \\ \text{s.t.} \quad & \mathbf{x}_j = \mathbf{X}\mathbf{z}_j, \\ & z_{jj} = 0, \end{aligned} \quad (4)$$

with

$$W_{i,j} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in N_K(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_K(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$S_{i,j} = 1 - W_{i,j},$$

where $N_K(\mathbf{x}_i)$ denotes the set of K -nearest neighbors of \mathbf{x}_i and W is the K -nn graph of the dataset. The first term $\sum_i S_{ij} |Z_{ij}|$ only minimizes the coefficients corresponding to the data points that are not connected. The second term enforces close representation vectors for connected close data. These two terms enforce the representation matrix and the resulting affinity matrix to have both sparsity and GE. The sparsity encourages grouping data from different subspaces into different clusters and the GE encourages grouping highly correlated data together.

Considering that $(1/2) \sum_{i,j} W_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 = \text{tr}(\mathbf{Z}^T \mathbf{L} \mathbf{Z})$, where $\mathbf{L} = \mathbf{D} - \mathbf{W}$, and \mathbf{D} is a diagonal matrix with diagonal entries $D_{ii} = \sum_j W_{ij}$ and collecting problem (5) for all data points, we have the following problem:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{C}, \mathbf{E}, \mathbf{Y}_1, \mathbf{Y}_2} \quad & L(\mathbf{Z}, \mathbf{C}, \mathbf{E}, \mathbf{Y}_1, \mathbf{Y}_2) \\ & = \|\mathbf{S} \odot \mathbf{Z}\|_1 + \lambda \text{tr}(\mathbf{C}^T \mathbf{L} \mathbf{C}) + \beta \|\mathbf{E}\|_l + \frac{\mu}{2} \left(\left\| \mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E} + \frac{\mathbf{Y}_1}{\mu} \right\|_F^2 + \left\| \mathbf{C} - \mathbf{Z} + \text{diag}(\mathbf{Z}) + \frac{\mathbf{Y}_2}{\mu} \right\|_F^2 \right), \end{aligned} \quad (9)$$

where \mathbf{Y}_1 and \mathbf{Y}_2 are matrices of Lagrange multipliers and $\mu > 0$ is a parameter. To find a saddle point for L , we update each of \mathbf{Z} , \mathbf{C} , \mathbf{E} , \mathbf{Y}_1 , and \mathbf{Y}_2 alternatively while keeping the other variables fixed.

Update for Z. We update \mathbf{Z} as follows:

$$\mathbf{Z}^{t+1} = \arg \min_{\mathbf{Z}} \frac{1}{\mu_t} \|\mathbf{S} \odot \mathbf{Z}\|_1 + \frac{1}{2} \left\| \mathbf{Z} - \text{diag}(\mathbf{Z}) - \mathbf{U}^t \right\|_F^2, \quad (10)$$

where $\mathbf{U}^t = \mathbf{C}^t + \mathbf{Y}_2^t / \mu_t$; the solution for (10) is

$$\mathbf{Z}^{t+1} = \bar{\mathbf{Z}} - \text{diag}(\bar{\mathbf{Z}}) \quad (11)$$

and $\bar{\mathbf{Z}}$ can be solved by using the soft-thresholding operator:

$$\bar{Z}_{ij} = \text{sgn}(U_{ij}^t) \max\left(|U_{ij}^t| - \frac{S_{ij}}{\mu_t}, 0\right). \quad (12)$$

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{S} \odot \mathbf{Z}\|_1 + \lambda \text{tr}(\mathbf{Z}^T \mathbf{L} \mathbf{Z}), \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{X}\mathbf{Z}, \\ & \text{diag}(\mathbf{Z}) = 0, \end{aligned} \quad (6)$$

where \odot denotes the element-wise product of two matrices. In case of noisy or corrupted data, we reformulate the problem as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \quad & \|\mathbf{S} \odot \mathbf{Z}\|_1 + \lambda \text{tr}(\mathbf{Z}^T \mathbf{L} \mathbf{Z}) + \beta \|\mathbf{E}\|_l \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \\ & \text{diag}(\mathbf{Z}) = 0, \end{aligned} \quad (7)$$

where the parameters $\lambda > 0$ and $\beta > 0$ are tuning parameters to balance the effect of the corresponding terms. The constraint $\text{diag}(\mathbf{Z}) = 0$ enforces that each data point be represented as a linear combination of other data points.

2.2. Minimization Algorithm. To solve problem (7), we first convert it into the following equivalent problem:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{C}, \mathbf{E}} \quad & \|\mathbf{S} \odot \mathbf{Z}\|_1 + \lambda \text{tr}(\mathbf{C}^T \mathbf{L} \mathbf{C}) + \beta \|\mathbf{E}\|_l \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{X}\mathbf{C} + \mathbf{E}, \\ & \mathbf{C} = \mathbf{Z} - \text{diag}(\mathbf{Z}). \end{aligned} \quad (8)$$

We solve this problem by using the Alternating Direction Method of Multipliers (ADMM) [16]. The augmented Lagrangian [17] is given by

Update for C. We update \mathbf{C} as follows:

$$\begin{aligned} \mathbf{C}^{t+1} = \arg \min_{\mathbf{C}} \quad & \lambda \text{tr}(\mathbf{C}^T \mathbf{L} \mathbf{C}) \\ & + \frac{\mu_t}{2} \left(\left\| \mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}^t + \frac{\mathbf{Y}_1^t}{\mu_t} \right\|_F^2 \right. \\ & \left. + \left\| \mathbf{C} - \mathbf{Z}^t + \text{diag}(\mathbf{Z}^t) + \frac{\mathbf{Y}_2^t}{\mu_t} \right\|_F^2 \right). \end{aligned} \quad (13)$$

This is a smooth convex program and the optimal solution condition is

$$\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{B} = \mathbf{D}, \quad (14)$$

Input: data matrix \mathbf{X} , model parameters λ and β and K .

Initialization: $\mathbf{Z}^0 = \mathbf{C}^0 = \mathbf{Y}_1^0 = \mathbf{Y}_2^0 = 0$, $\mu_0 = 0.1$,

$\mu_{\max} = 10^{10}$, $t = 0$, $\rho = 1.1$, $\varepsilon = 10^{-5}$.

Output: the optimal coefficient matrix \mathbf{Z}^* .

While not converge **do**

(1) fix others and update \mathbf{Z} by Eq. (11) and (12);

(2) fix others and update \mathbf{C} by solving Eq. (14);

(3) fix others and update \mathbf{E} by Eq. (16) or (17);

(4) update the Lagrangian Multipliers:

$$\mathbf{Y}_1^{t+1} = \mathbf{Y}_1^t + \mu_t(\mathbf{X} - \mathbf{X}\mathbf{C}^t - \mathbf{E}^t)$$

$$\mathbf{Y}_2^{t+1} = \mathbf{Y}_2^t + \mu_t(\mathbf{C}^t - \mathbf{Z}^t + \text{diag}(\mathbf{Z}^t))$$

(5) update the parameter μ by $\mu_{t+1} = \min(\mu_{\max}, \rho\mu_t)$;

(6) check the convergence conditions

$$\|\mathbf{X} - \mathbf{X}\mathbf{C}_t - \mathbf{E}_t\|_{\infty} < \varepsilon$$

$$\|\mathbf{C}_t - \mathbf{Z}_t + \text{diag}(\mathbf{Z}_t)\|_{\infty} < \varepsilon$$

(7) $t = t + 1$;

end while

ALGORITHM 1: Solving problem (7) by ADMM.

where $\mathbf{A} = \mu_t(\mathbf{X}^T\mathbf{X} + \mathbf{I})$, $\mathbf{D} = \mu_t(\mathbf{X}^T\mathbf{X} - \mathbf{X}^T\mathbf{E}^t + \mathbf{Z}^t - \text{diag}(\mathbf{Z}^t)) + \mathbf{X}^T\mathbf{Y}_1^t - \mathbf{Y}_2^t$, and $\mathbf{B} = 2\lambda\mathbf{L}$. Equation (14) is a standard Sylvester equation [18], which has one unique solution.

Update for E. We update \mathbf{E} as follows:

$$\mathbf{E}^{t+1} = \arg \min_{\mathbf{E}} \beta \|\mathbf{E}\|_l + \frac{\mu_t}{2} \left\| \mathbf{X} - \mathbf{X}\mathbf{C}^t - \mathbf{E} + \frac{\mathbf{Y}_1^t}{\mu_t} \right\|_F^2. \quad (15)$$

Here we use l_1 -norm or squared Frobenius norm to handle corruptions or noise. If l_1 -norm is used, the solution for (15) is

$$\begin{aligned} \mathbf{E}_{ij}^{t+1} &= \text{sgn} \left(\left(\mathbf{X} - \mathbf{X}\mathbf{C}^t + \frac{\mathbf{Y}_1^t}{\mu_t} \right)_{ij} \right) \\ &\cdot \max \left(\left| \left(\mathbf{X} - \mathbf{X}\mathbf{C}^t + \frac{\mathbf{Y}_1^t}{\mu_t} \right)_{ij} \right| - \frac{\beta}{\mu_t}, 0 \right). \end{aligned} \quad (16)$$

If squared Frobenius norm is used, the solution is

$$\mathbf{E}^{t+1} = (2\beta + \mu_t)^{-1} (\mu_t\mathbf{X} - \mu_t\mathbf{X}\mathbf{C}^t + \mathbf{Y}_1^t). \quad (17)$$

Update for \mathbf{Y}_1 and \mathbf{Y}_2 . We update \mathbf{Y}_1 and \mathbf{Y}_2 as follows:

$$\begin{aligned} \mathbf{Y}_1^{t+1} &= \mathbf{Y}_1^t + \mu_t(\mathbf{X} - \mathbf{X}\mathbf{C}^t - \mathbf{E}^t) \\ \mathbf{Y}_2^{t+1} &= \mathbf{Y}_2^t + \mu_t(\mathbf{C}^t - \mathbf{Z}^t + \text{diag}(\mathbf{Z}^t)). \end{aligned} \quad (18)$$

The update for the Lagrange multipliers is a simple gradient descent step.

The procedure for solving problem (7) is outlined in Algorithm 1. When the optimal solution \mathbf{Z}^* of problem (7) is obtained, we use (2) to construct the affinity matrix \mathbf{A} and normalize each column vector of \mathbf{A} by dividing

the maximum value of that column. Finally, the spectral clustering method is applied to obtain the clustering results. The computation burden is mainly focused on Algorithm 1. Let k denote the number of iterations and let the matrix \mathbf{X} have size $d \times n$. The complexity of updating \mathbf{Z} , \mathbf{E} , and the multipliers \mathbf{Y}_1 and \mathbf{Y}_2 is, respectively, $O(n^2)$, $O(dn^2)$, $O(dn^2)$, and $O(n^2)$. In the process of updating \mathbf{C} , we used the Bartels-Stewart algorithm [18] to solve the Sylvester equation; the algorithm has a computational complexity of $O(n^3)$. So the total computational complexity is $O(kn^2 + kdn^2 + kn^3)$. Assuming that $d \leq n$, the complexity is $O(kn^3)$.

3. Experiments and Analysis

In this section, we evaluate the performance of the proposed method on several publicly available datasets, including Extended Yale B dataset [14], the handwritten digits dataset USPS [15], and Berkeley 500 image segmentation dataset [19]. We also compare our method with several state-of-the-art methods: SSC [7], LRR [2], LSR [9], CASS [10], and SMR [11]. We use the source codes provided by the authors and tune the parameters of each method to achieve the best performance for fair comparison. We define the affinity matrix by using (2).

For the Extended Yale B and USPS datasets, we use the subspace clustering error to evaluate the performance of the methods:

$$\text{error} = \frac{N_{\text{error}}}{N_{\text{total}}}, \quad (19)$$

where N_{error} denotes the number of misclassified points and N_{total} denotes the total number of points. For the Berkeley 500 image segmentation dataset, three metrics for comparing pairs of image segmentations are used: PRI (probabilistic rand index) [20], VOI (variation of information) [21], and GCE (global consistency error) [22]. The value of PRI is within $[0, 1]$ and the higher value shows the better segmentation performance. The value of VOI is within $[0, +\infty)$ and the lower value shows the better segmentation performance. The value of GCE is within $[0, 1]$ and the lower value shows the better performance.

3.1. Datasets and Experimental Settings

Extended Yale B. Extended Yale B is a face dataset that consists of 192×168 -pixel cropped face images of 38 human subjects under varying poses and illuminations. Each subject has 64 frontal faces. To reduce the computational cost and memory requirements of all algorithms, each face image is downsampled to 32×32 pixels and rearranged into a 1024-dimensional vector. To study the effect of the number of subjects on the clustering performance of different methods, we divide the 38 subjects into four groups: 1 to 10, 11 to 20, 21 to 30, and 31 to 38; for the first three groups (each group has 10 subjects in all), we test all methods on 2, 3, 5, 8, and 10 subjects; for the last group (has 8 subjects), we perform experiments on 2, 3, 5, and 8 subjects. Finally, the mean and median clustering errors of different number of subjects are reported for each method to evaluate their performance.

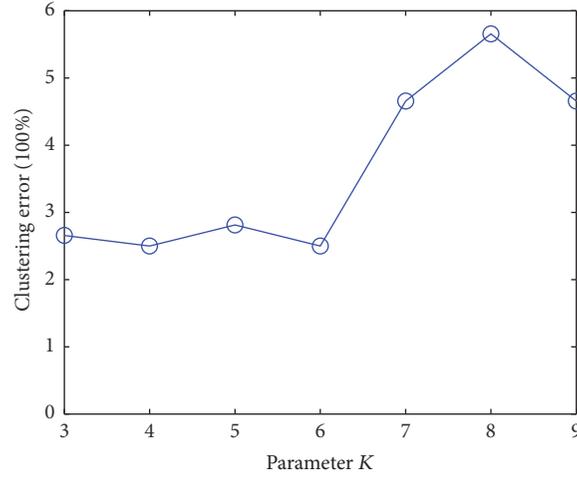


FIGURE 2: The clustering error varies with the size of neighborhood K .

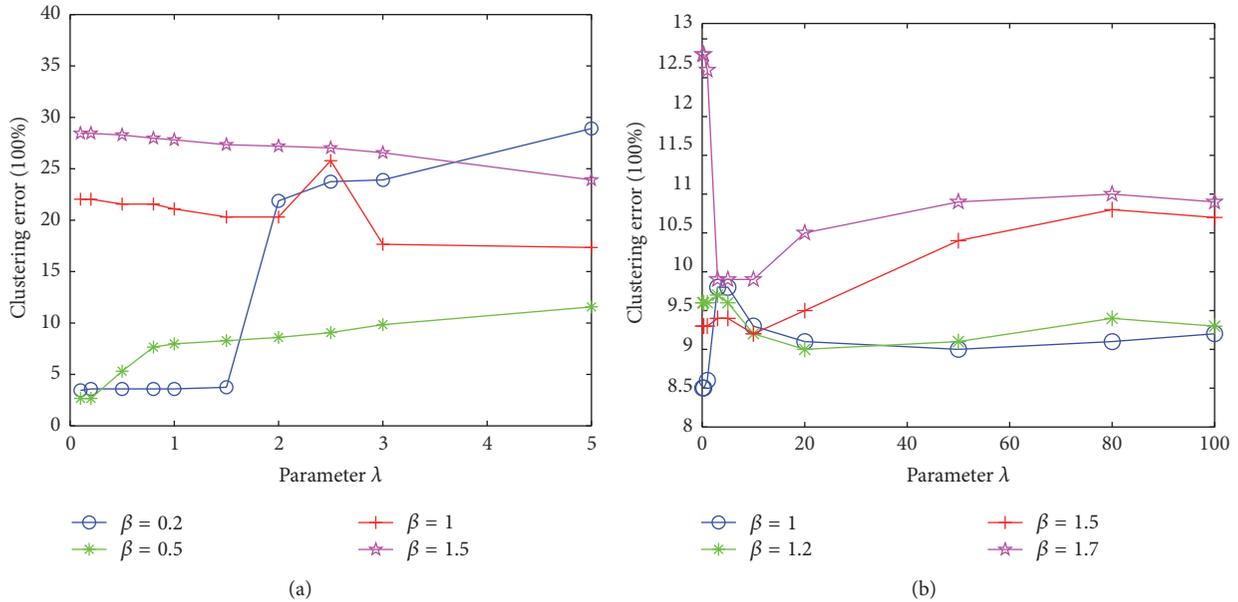


FIGURE 3: The clustering error rate varies with the parameters λ and β on Extend Yale B dataset (a) and USPS dataset (b).

USPS. USPS is a database of handwritten digits. It consists of 10 subjects, corresponding to 10 handwritten digits, 0–9. There are 9298 images, with each image having 16×16 pixels. We use the first 100 images of each digit in experiment.

Berkeley 500. Berkeley 500 is an image dataset consisting of 500 multiple natural images for segmentation. It covers a variety of scene categories in nature, such as portraits, animals, landscape, and beaches. It also provides ground truth segmentation results of all the images obtained by several human subjects. We select 50 images from this dataset in experiment for image segmentation. We first partition each image into superpixels using the method presented in [23]. Then we extract the Color Histogram (CH), Local Binary Pattern (LBP), and Histogram of Gradient (HOG) features of each superpixel and obtained the data X . Finally, we use our subspace clustering method to cluster the superpixels into several regions.

3.2. Parameter Setting. There are three model parameters, λ , β , and K , affecting the performance of our method. In the following, we analyze the effect of these three parameters on the clustering performance. We choose the first 10 subjects of Extend Yale B and USPS as the test database. We first test the influence of K . Figure 2 shows the clustering error rate varying with the size of neighborhood K from 3 to 9 on Extend Yale B database, with $\lambda = 0.1$ and $\beta = 0.5$. It shows that our method performs best with $K = 6$. On USPS, we have the same result. So we take $K = 6$ in all experiments.

Figure 3 shows how the clustering error rate of the proposed method depends on the parameter λ , with several fixed values of β on the datasets Extend Yale B (a) and USPS (b). We applied our method on the first 10 subjects of Extend Yale B and all subjects of USPS. For Extend Yale B, one can observe that for $\beta = 0.2$ and $\beta = 0.5$ the proposed method

TABLE 1: The clustering error rate (%) of Extended Yale B dataset.

	SSC	LRR	LSR	CASS	SMR	Ours
2 subjects						
Mean	0.00	4.08	3.53	0.00	4.87	0.00
Median	0.00	3.91	2.23	0.00	3.91	0.00
3 subjects						
Mean	1.43	6.51	6.05	0.69	6.70	1.30
Median	0.78	4.69	4.69	0.00	6.25	0.26
5 subjects						
Mean	3.15	9.04	6.95	2.02	14.00	2.06
Median	1.92	4.93	4.28	2.30	14.69	1.62
8 subjects						
Mean	6.93	18.24	24.21	18.36	24.91	3.26
Median	4.20	30.47	21.86	16.40	24.90	3.58
10 subjects						
Mean	10.29	35.89	30.66	22.65	29.38	2.91
Median	10.52	35.62	30.47	21.54	29.22	2.78

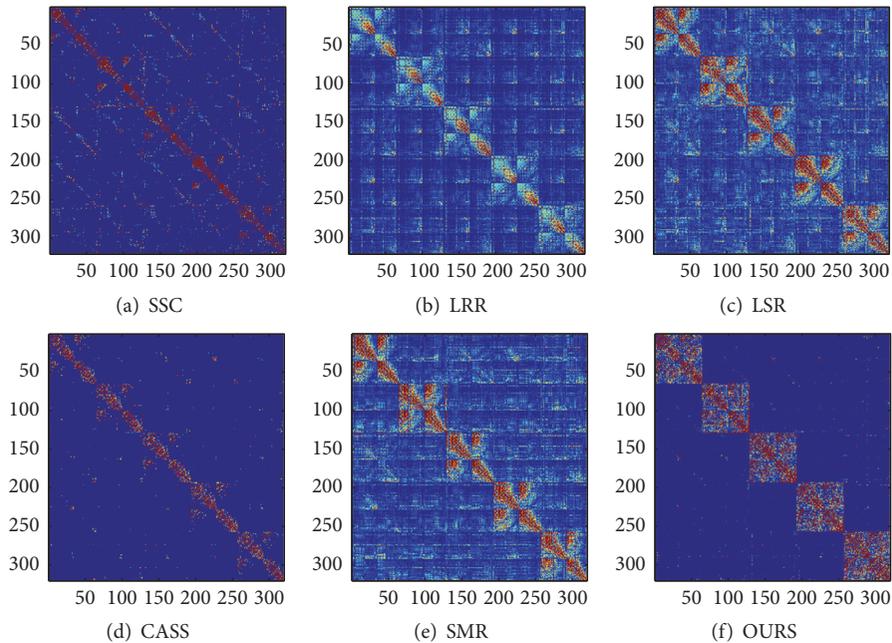


FIGURE 4: Affinity matrices of 5 subjects from Extend Yale B dataset by different methods.

performs well with $\lambda \in [0, 1.5]$. We set $\beta = 0.5$ because the error rate behaves more stably with respect to λ . So we set $\beta = 0.5$ and $\lambda = 0.1$ because the error rate obtains the smallest value in this case. For USPS, the clustering error rate behaves relatively stably with respect to λ when $\beta = 1$ or $\beta = 1.2$. We set $\beta = 1$ and $\lambda = 0.1$ because this leads to the smallest error rate.

3.3. Experimental Results. It should be noted that the Extend Yale B dataset is challenging for subspace clustering due to heavy corruptions in the data. To show the effectiveness, we select all the face images in test. We use the l_1 -norm to measure the representation error matrix \mathbf{E} , and we set

$\lambda = 0.1$, $\beta = 0.5$, and $K = 6$. We divide the 38 subjects into four groups: 1 to 10, 11 to 20, 21 to 30, and 31 to 38; for the first three groups (each group has 10 subjects in all), we test all methods on 2, 3, 5, 8, and 10 subjects; for the last group (has 8 subjects), we perform experiments on 2, 3, 5, and 8 subjects. Finally, we calculate the mean and median clustering error rates of the four groups. The clustering error rates by all methods in comparison are presented in Table 1. The shapes of affinity matrices of 5 subjects obtained by all methods are illustrated in Figure 4. Our method obtains an affinity matrix which is most close to block diagonal and obtains the lowest error rate in case of multiple subjects. The clustering error rates of LRR, LSR, CASS, and SMR increase

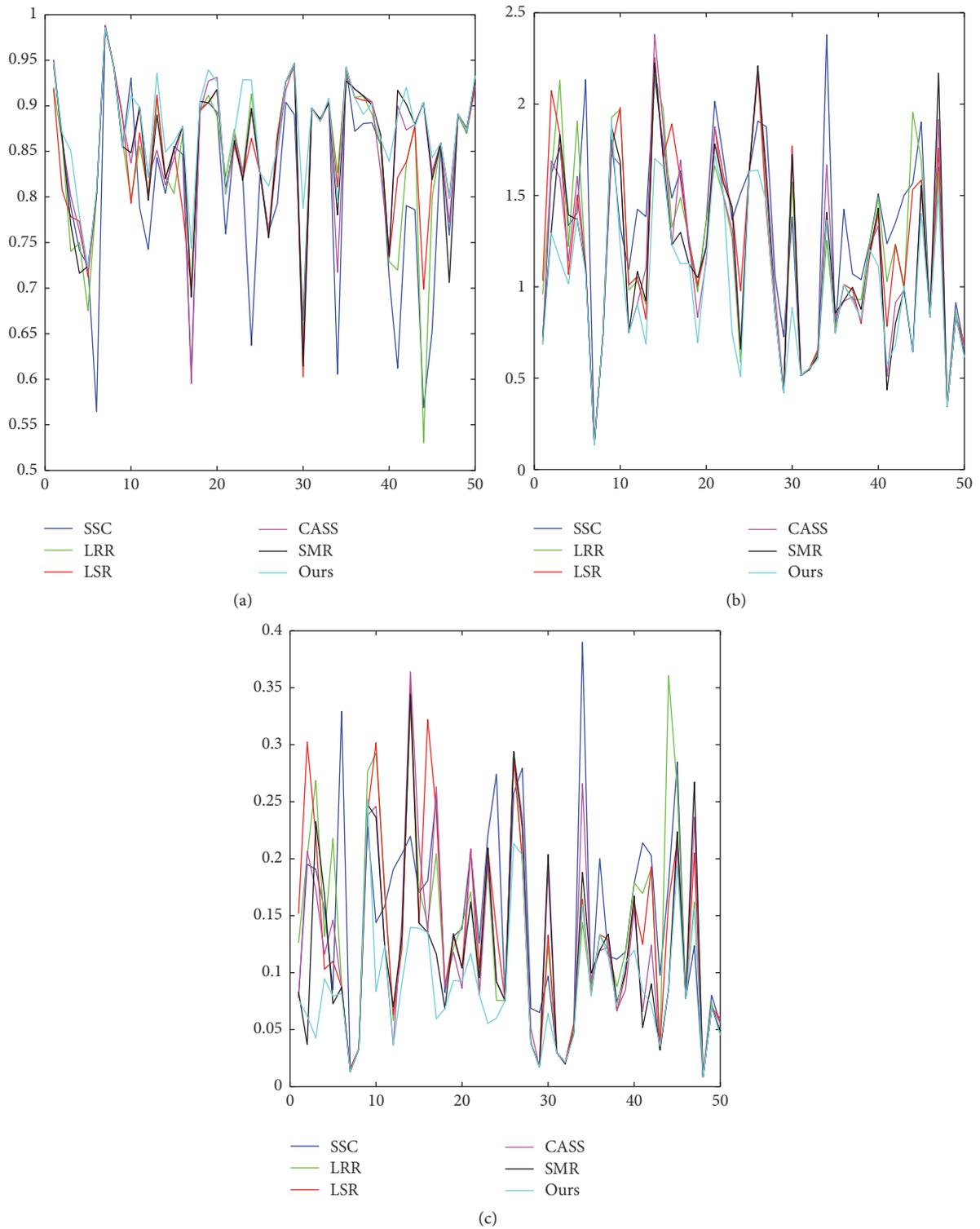


FIGURE 5: The values of the three metrics, PRI (higher is better), VOI (lower is better), and GCE (lower is better), on the 50 images produced by SSC, LRR, LSR, CASS, SMR, and our method. (a) PRI, (b) VOI, and (c) GCE.

quickly when the number of subjects increases. This is mainly because their affinity matrices are not strictly block diagonal, thus leading to grouping data from different subspaces into the same cluster. Due to the sparsity between clusters, the performance of SSC degenerates very slowly as the number of

subjects increases. But our method behaves most stably with respect to the number of subjects and it consistently obtains the lowest clustering error rate among all methods.

For USPS dataset, we perform the clustering experiment on all data points of all subjects. We use the Frobenius norm



FIGURE 6: Some segmentation results of the Berkeley 500 dataset by our method.

to measure the representation error matrix, and we set $\lambda = 0.1$, $\beta = 1$, and $K = 6$. Table 2 presents the clustering error rates, which shows that our method performs best and it reduces the clustering error rate of other methods greatly.

For the Berkeley 500 image segmentation dataset, we select 50 images randomly for image segmentation and use three metrics (PRI, VOI, and GCE) for objective assessment of the methods used. Because of the variety of structures underlying the natural images, we tune the parameters to achieve the best results. Table 3 shows the values of the three metrics averaged on the 50 images. We can see that our method can obtain a PRI of 0.8745, a VOI of 1.0068, and a GCE of 0.0886, which distinctly outperforms other methods. The detailed values of the three metrics on the 50 images are shown in Figure 5. Some examples of the segmentation results obtained by our method are shown in Figure 6 for visual assessment. It can be seen that the segmentation results produced by our method are similar to the ground truth and are visually satisfying.

In all, the experimental results on the three datasets show that, by taking both sparsity and grouping effect into consideration, our method is more efficient in recovering the subspace structure underlying high-dimensional dataset.

TABLE 2: The clustering error rate (%) of USPS dataset.

SSC	LRR	LSR	CASS	SMR	Ours
13.50	22.60	26.10	18.00	12.70	8.60

4. Conclusion

The state-of-the-art methods show that sparsity and grouping effect of the affinity matrix are useful in clustering. In this work, we enforce both requirements in modeling the self-representation of data points. The experimental results on Extended Yale B, USPS, and Berkeley 500 image segmentation datasets show the requirements of sparsity and grouping effect and the proposed model is more effective than state-of-the-art methods in revealing the cluster structure underlying high-dimensional dataset.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank the National Natural Science Foundation of China (Grants no. 61472303 and no. 61271294)

TABLE 3: The average metrics of PRI, VOI, and GCE on Berkeley 500 image segmentation dataset.

	SSC	LRR	LSR	CASS	SMR	Ours
PRI	0.8118	0.8410	0.8460	0.8531	0.8546	0.8745
VOI	1.3151	1.2166	1.2205	1.1557	1.1449	1.0068
GCE	0.1495	0.1369	0.1360	0.1265	0.1186	0.0886

and the Fundamental Research Funds for the Central Universities (Grant no. NSIY21) for supporting their research works.

References

- [1] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data: a review," *SIGKDD Explorations*, vol. 6, no. 1, pp. 90–105, 2004.
- [2] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.
- [3] K. Kanatani, "Motion segmentation by subspace separation and model selection," in *Proceedings of the 8th International Conference on Computer Vision*, pp. 586–591, Vancouver, Canada, July 2001.
- [4] B. Cheng, G. Liu, J. Wang, Z. Huang, and S. Yan, "Multi-task low-rank affinity pursuit for image segmentation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2439–2446, IEEE, Barcelona, Spain, November 2011.
- [5] C. Zhang and R. R. Bitmead, "Subspace system identification for training-based MIMO channel estimation," *Automatica*, vol. 41, no. 9, pp. 1623–1632, 2005.
- [6] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2790–2797, June 2009.
- [7] E. Elhamifar and R. Vidal, "Sparse subspace clustering: algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.
- [8] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the 27th International Conference on Machine Learning (ICML '10)*, pp. 663–670, June 2010.
- [9] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *Proceedings of the IEEE 12th European Conference on Computer Vision*, pp. 347–360, Florence, Italy, October 2012.
- [10] C. Lu, J. Feng, Z. Lin, and S. Yan, "Correlation adaptive subspace segmentation by trace lasso," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 1345–1352, IEEE, Sydney, Australia, December 2013.
- [11] H. Hu, Z. Lin, J. Feng, and J. Zhou, "Smooth representation clustering," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 3834–3841, Columbus, Ohio, USA, June 2014.
- [12] D.-S. Pham, S. Budhaditya, D. Phung, and S. Venkatesh, "Improved subspace clustering via exploitation of spatial constraints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 550–557, June 2012.
- [13] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [14] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [15] J. J. Hull, "A database for handwritten text recognition research," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 550–554, 1994.
- [16] X. Yuan and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," *Pacific Journal of Optimization*, vol. 9, no. 1, 2009.
- [17] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2009.
- [18] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation $AX + XB = C$ [F4]," *Communications of the ACM*, vol. 15, no. 9, pp. 820–826, 1972.
- [19] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.
- [20] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.
- [21] M. Meila, "Comparing clusterings: an axiomatic view," in *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, pp. 577–584, Bonn, Germany, August 2005.
- [22] D. Zhou, O. Bousquet, and T. Lal, "Learning with local and global consistency," *Advances in Neural Information Processing Systems*, vol. 16, no. 16, pp. 321–328, 2004.
- [23] G. Mori, X. Ren, A. A. Efros, and J. Malik, "Recovering human body configurations: combining segmentation and recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. II-326–II-333, USA, July 2004.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

