



Robust Subspace Segmentation via Sparse Relation Representation

Lai Wei¹(✉)  and Hao Liu²

¹ Shanghai Maritime University,
Haigang Avenue 1550, Shanghai, People's Republic of China
weilai@shmtu.edu.cn

² Wuhan Digit Engineering Institute,
Canglong North Road 709, Wuhan, Hubei, People's Republic of China
sjupiter83@hotmail.com

Abstract. Spectral clustering based algorithms are powerful tools for solving subspace segmentation problems. The existing spectral clustering based subspace segmentation algorithms use original data matrices to produce the affinity graphs. In real applications, data samples are usually corrupted by different kinds of noise, hence the obtained affinity graphs may not reveal the intrinsic subspace structures of data sets. In this paper, we present the conception of relation representation, which means a point's neighborhood relation could be represented by the rest points' neighborhood relations. Based on this conception, we propose a kind of sparse relation representation (SRR) for subspace segmentation. The experimental results obtained on several benchmark databases show that SRR outperforms some existing related methods.

Keywords: Subspace segmentation · Low-rank representation · Sparse relation

1 Introduction

Spectral clustering based algorithms have been proven to be powerful tools for solving subspace segmentation problems such as motion segmentation [1, 2], face clustering [3, 4] and so on. Among the existing spectral clustering based methods, sparse subspace clustering (SSC) [5] and low-rank representation (LRR) [6] are the two most representative ones. The two algorithms divide the subspace segmentation procedure into three steps: firstly, they compute a reconstructive coefficient matrix for a data set, then construct an affinity graph by using the obtained coefficient matrix, finally produce the segmentation result by means of a kind of spectral clustering (e.g. Normalize cut (N-cut) [7]). Because of the excellent performances showed by SSC and LRR, a lot of subsequent researches have been proposed.

By analyzing SSC and LRR related works, we could find that most of them hope to enhance the abilities of SSC and LRR for revealing subspace structures

of data sets by adding additional constraints on the reconstructive coefficient matrices. For example, Li et al. devised an adaptive weighted sparse constraint for a reconstructive coefficient matrix obtained by SSC and proposed a structured SSC method (SSSC) [8]. Chen et al. developed a within-class grouping constraint for a reconstructive coefficient matrix and introduced it into SSSC [9]. Zhuang et al. claimed that SSC tends to discover the local structure of a data set and LRR could discover its global structure. Hence, they combined SSC and LRR together and proposed a non-negative low-rank and sparse representation method (NNLRSR) [10]. Tang et al. generalized NNLRSR algorithm and designed a structured-constrained LRR method (SCLRR) [11]. Lu et al. presented a graph-regularized low-rank representation (GLRR) [12] which could strength the group effect of a coefficient matrix obtained by LRR.

According to the corresponding references, the above mentioned algorithms have shown to be superior to the classical SSC and LRR. However, we could find these algorithms follow the same methodology of SSC and LRR as we mentioned in the first paragraph.

In this paper, we reconsider the data representation problem and present the concept and technique of *relation representation*. Based on these new propositions, we propose a new algorithm, termed sparse relation representation (SRR), for subspace segmentation. We claim that SRR could find both the local and global structures of data sets. The experimental results obtained on different subspace segmentation tasks illustrate that SRR dominates the existing SSC and LRR related algorithms.

The rest of the paper is organized as follows: we briefly review SSC and LRR algorithms in Sect. 2. In Sect. 3, we introduce the idea of relation representation and present sparse relation representation (SRR) method. The optimization algorithm for solving SRR problem is described in Sect. 4. Experiments on benchmark data sets are conducted in Sect. 5. Section 6 gives the conclusions.

2 Preliminary

For a data set $\mathbf{X} \in \mathcal{R}^{d \times n}$, both SSC and LRR hope to find a reconstruction matrix $\mathbf{C} \in \mathcal{R}^{n \times n}$ which satisfies $\mathbf{X} = \mathbf{XC} + \mathbf{E}$. Here, $\mathbf{E} \in \mathcal{R}^{d \times n}$ indicates the reconstruction residual matrix. With different techniques, SSC expects \mathbf{C} to be a sparse matrix and the l_1 norm of \mathbf{E} to be minimal, while LRR tends to minimize the rank of \mathbf{C} and the $l_{2,1}$ norm of \mathbf{E} simultaneously. Because of the different constraints imposed on the coefficient matrix, SSC and LRR tends to reveal the local and global structures of a data set respectively. The objective function of SSC and LRR could be precisely expressed as the following Eqs. 1 and 2 respectively:

$$\begin{aligned} \min \quad & \|\mathbf{C}\|_1 + \lambda \|\mathbf{E}\|_1, \\ \text{s.t. } \quad & \mathbf{X} = \mathbf{XC} + \mathbf{E}, [\mathbf{C}]_{ii} = 0, i = 1, 2, \dots, n, \end{aligned} \quad (1)$$

$$\begin{aligned} \min \quad & \|\mathbf{C}\|_* + \lambda \|\mathbf{E}\|_{2,1}, \\ \text{s.t. } \quad & \mathbf{X} = \mathbf{XC} + \mathbf{E}, \end{aligned} \quad (2)$$

where $[C]_{ii}$ represents the (i, i) -th element of \mathbf{C} and $\lambda > 0$ is a parameter which is used to balance the effects of the two terms. The above two problems could be solved by using the alternating direction method (ADM) [13]. Once the reconstructive coefficient matrix \mathbf{C} is gotten, an affinity matrix \mathbf{W} satisfying $[W]_{ij} = ([C]_{ij} + [C]_{ji})/2$ could be constructed. Then the final segmentation result could be produced by using N-cut.

3 Motivation

3.1 Relation Representation

From the above descriptions, we could find that SSC and LRR (actually all the related algorithms) use a data set itself to compute the reconstruction coefficient matrix. However, in real applications, data samples usually is corrupted by different kinds of noise, so the obtained coefficient matrix may not be able to reveal the subspace structure of a data set.

As we know, the relationships between an object and its neighbors could usually define the object itself. And two similar objects will often have similar neighbors with similar relationships (See Fig. 1). Based on these evidences, for a data set, we consider to use the relations between a data sample and its neighbors to represent the data sample firstly, then reconstruct the neighborhood relation of a data sample by using the neighborhood relations of other samples. Hence, the reconstruction coefficient vector corresponding to each data sample's neighborhood relation could be acquired. We call this strategy "relation representation".

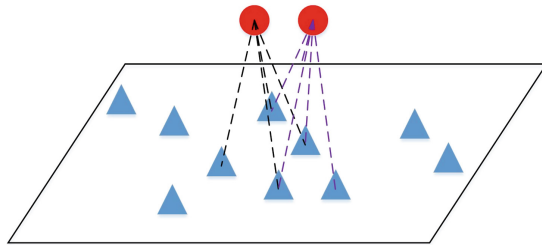


Fig. 1. Two similar objects (red points) and their neighbors (blue triangles) (Color figure online)

3.2 Sparse Relation Representation (SRR)

Now we discuss how to compute the relations between a data sample and its neighbors. Actually, many skills could handle this problem. For example, K nearest neighbors method (KNN) [14] can find each data sample's K neighbors,

then linear reconstruction method [15] or Gaussian kernel [14] could be used to compute the similarities between the data sample and its K neighbors. However, the neighborhood scale K in KNN is usually difficult to choose for different data sets. And an improper K will degenerate the performance of corresponding algorithm sharply.

It has been proven that sparse representation (SR) technique [16] is capable of adaptively choosing the neighbors of a data sample and getting the corresponding reconstruction coefficient simultaneously. Therefore, for a data sample $\mathbf{x}_i \in \mathbf{X}$, its neighborhood relation vector \mathbf{c}_i could be achieved by using the following SR problem:

$$\min_{\mathbf{c}_i} \|\mathbf{c}_i\|_1 + \alpha \|\mathbf{x}_i - \mathbf{X}\mathbf{c}_i\|_1. \quad (3)$$

We hope the reconstruction residual $\mathbf{x}_i - \mathbf{X}\mathbf{c}_i$ also to be sparse. Then for the whole data matrix \mathbf{X} , we could get its neighborhood relation matrix \mathbf{C} by solving the following problem:

$$\begin{aligned} \min_{\mathbf{C}} \|\mathbf{C}\|_1 + \alpha \|\mathbf{X} - \mathbf{X}\mathbf{C}\|_1, \\ \text{s.t. } [\mathbf{C}]_{ii} = 0, i = 1, 2, \dots, n \end{aligned} \quad (4)$$

Similar to SSC, we could find that \mathbf{C} will discover the local structure of the original data set.

Then according to the relation representation technique (described in Sect. 3.1), a data sample \mathbf{x}_i 's neighborhood relation \mathbf{c}_i could be represented by the neighborhood relations of other data samples, namely $\mathbf{c}_i \simeq \mathbf{C}\mathbf{z}_i$, where $\mathbf{z}_i \in \mathcal{R}^n$ is the reconstruction coefficient. Consider the whole data set, we could obtain the following problem:

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_* + \beta \|\mathbf{C} - \mathbf{C}\mathbf{Z}\|_F^2 \quad (5)$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n]$ is the reconstruction coefficient matrix to the neighborhood relation matrix \mathbf{C} . We here use the nuclear norm minimization constraint to help \mathbf{Z} to discover the global structure of \mathbf{C} . Moreover, Because \mathbf{C} is a good representation of original data matrix \mathbf{X} , we aim to minimize the Frobenius norm of the reconstruction error. Finally, we combine Eqs. 4 and 5 together and let $\mathbf{E}_1 = \mathbf{X} - \mathbf{X}\mathbf{C}$, $\mathbf{E}_2 = \mathbf{C} - \mathbf{C}\mathbf{Z}$, then the sparse relation representation problem (SRR) could be defined as follows:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{Z}, \mathbf{E}_1, \mathbf{E}_2} \|\mathbf{C}\|_1 + \|\mathbf{Z}\|_* + \alpha \|\mathbf{E}_1\|_1 + \beta \|\mathbf{E}_2\|_F^2, \\ \text{s.t. } \mathbf{E}_1 = \mathbf{X} - \mathbf{X}\mathbf{C}, \\ \mathbf{E}_2 = \mathbf{C} - \mathbf{C}\mathbf{Z}, \\ [\mathbf{C}]_{ii} = 0, i = 1, 2, \dots, n. \end{aligned} \quad (6)$$

For a data set \mathbf{X} , because \mathbf{C} characterizes the local structure of \mathbf{X} and \mathbf{Z} discovers the global structure of the neighborhood relation matrix \mathbf{C} , \mathbf{Z} actually could reveal both the global and local structure of a data set.

4 Optimization and Analyses

4.1 Optimization

For solving Problem 6, we firstly covert it into the following equivalent problem:

$$\begin{aligned}
 & \min_{\mathbf{C}, \mathbf{Z}, \mathbf{M}, \mathbf{J}, \mathbf{E}_1, \mathbf{E}_2} \|\mathbf{M}\|_1 + \|\mathbf{J}\|_* + \alpha \|\mathbf{E}_1\|_1 + \beta \|\mathbf{E}_2\|_F^2, \\
 & \text{s.t.} \quad \mathbf{E}_1 = \mathbf{X} - \mathbf{X}\mathbf{C}, \\
 & \quad \mathbf{C} = \mathbf{M}, [\mathbf{M}]_{ii} = 0, i = 1, 2, \dots, n, \\
 & \quad \mathbf{E}_2 = \mathbf{C} - \mathbf{C}\mathbf{Z}, \\
 & \quad \mathbf{Z} = \mathbf{J}.
 \end{aligned} \tag{7}$$

The above could be solved by using ADM method [13]. The augmented Lagrangian function of Eq. 7 can be described as follows:

$$\begin{aligned}
 \mathcal{L} = & \|\mathbf{M}\|_1 + \|\mathbf{J}\|_* + \alpha \|\mathbf{E}_1\|_1 + \beta \|\mathbf{E}_2\|_F^2 + \langle \mathbf{Y}_1, \mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}_1 \rangle + \langle \mathbf{Y}_2, \mathbf{C} - \mathbf{M} \rangle \\
 & + \langle \mathbf{Y}_3, \mathbf{C} - \mathbf{C}\mathbf{Z} - \mathbf{E}_2 \rangle + \langle \mathbf{Y}_4, \mathbf{Z} - \mathbf{J} \rangle + \frac{\mu}{2} \left(\|\mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}_1\|_F^2 + \|\mathbf{C} - \mathbf{M}\|_F^2 \right. \\
 & \left. + \|\mathbf{C} - \mathbf{C}\mathbf{Z} - \mathbf{E}_2\|_F^2 + \|\mathbf{Z} - \mathbf{J}\|_F^2 \right),
 \end{aligned} \tag{8}$$

where $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$ and \mathbf{Y}_4 are four Lagrange multipliers, $\mu > 0$ is a parameter. Then by minimizing \mathcal{L} , the variables $\mathbf{C}, \mathbf{Z}, \mathbf{M}, \mathbf{J}, \mathbf{E}_1, \mathbf{E}_2$ could be optimized alternately. The detailed updating process for each variables presented as follows:

1. **Update \mathbf{M} with fixed other variables.** By collecting the relevant terms of \mathbf{M} in Eq. 8, we have:

$$\begin{aligned}
 & \min_{\mathbf{M}} \|\mathbf{M}\|_1 + \langle \mathbf{Y}_2, \mathbf{C} - \mathbf{M} \rangle + \mu/2 \|\mathbf{C} - \mathbf{M}\|_F^2 \\
 & = \min_{\mathbf{M}} \|\mathbf{M}\|_1 + \mu/2 \|\mathbf{C} - \mathbf{M} + \mathbf{Y}_2/\mu\|_F^2,
 \end{aligned} \tag{9}$$

then the solution to Eq. 9 could be obtained as

$$[\mathbf{M}^{opt}]_{ij} = \begin{cases} \max(0, [\mathbf{C} + \mathbf{Y}_2/\mu]_{ij} - 1/\mu) + \min(0, [\mathbf{C} + \mathbf{Y}_2/\mu]_{ij} + 1/\mu), & i \neq j; \\ 0, & i = j. \end{cases} \tag{10}$$

2. **Update \mathbf{C} with fixed other variables.** By picking the relevant terms of \mathbf{C} in Eq. 8, we have:

$$\begin{aligned}
 & \min_{\mathbf{C}} \langle \mathbf{Y}_1, \mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}_1 \rangle + \langle \mathbf{Y}_2, \mathbf{C} - \mathbf{M} \rangle + \langle \mathbf{Y}_3, \mathbf{C} - \mathbf{C}\mathbf{Z} - \mathbf{E}_2 \rangle \\
 & + \frac{\mu}{2} \left(\|\mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}_1\|_F^2 + \|\mathbf{C} - \mathbf{M}\|_F^2 + \|\mathbf{C} - \mathbf{C}\mathbf{Z} - \mathbf{E}_2\|_F^2 \right) \\
 & = \min_{\mathbf{C}} \|\mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}_1 + \mathbf{Y}_1/\mu\|_F^2 + \|\mathbf{C} - \mathbf{M} + \mathbf{Y}_2/\mu\|_F^2 \\
 & + \|\mathbf{C} - \mathbf{C}\mathbf{Z} - \mathbf{E}_2 + \mathbf{Y}_3/\mu\|_F^2
 \end{aligned} \tag{11}$$

We take the derivation of Eq. 11 w.r.t. \mathbf{C} and set it to $\mathbf{0}$, the following equation holds:

$$\begin{aligned}
 & (\mathbf{X}^t \mathbf{X} + \mathbf{I}_n) \mathbf{C}^{opt} + \mathbf{C}^{opt} (\mathbf{I}_n - \mathbf{Z}) (\mathbf{I}_n - \mathbf{Z}^t) - \mathbf{X}^t (\mathbf{X} - \mathbf{E}_1 + \mathbf{Y}_1/\mu) \\
 & - \mathbf{M} + \mathbf{Y}_2/\mu - (\mathbf{E}_2 - \mathbf{Y}_3/\mu) (\mathbf{I}_n - \mathbf{Z}^t) = \mathbf{0},
 \end{aligned} \tag{12}$$

where \mathbf{I}_n is an $n \times n$ identity matrix and \mathbf{X}^t and \mathbf{Z}^t are the transposes of \mathbf{X} and \mathbf{Z} respectively. Equation 12 is a Sylvester equation w.r.t. \mathbf{C}^{opt} , so it can be solved by the Matlab function `lyap()`.

3. **Update \mathbf{J} with fixed other variables.** By abandoning the irrelevant terms of \mathbf{J} , minimizing Eq. 8 becomes to the following problem:

$$\begin{aligned} & \min_{\mathbf{J}} \|\mathbf{J}\|_* + \langle \mathbf{Y}_4, \mathbf{Z} - \mathbf{J} \rangle + \mu/2 \|\mathbf{Z} - \mathbf{J}\|_F^2 \\ & = \min_{\mathbf{J}} \|\mathbf{J}\|_* + \mu/2 \|\mathbf{Z} - \mathbf{J} + \mathbf{Y}_4/\mu\|_F^2. \end{aligned} \quad (13)$$

Then the optimal solution to Eq. 13, $\mathbf{J}^{opt} = \mathbf{U}\Theta_{1/\mu}(\mathbf{S})\mathbf{V}$, where \mathbf{USV} is the SVD of matrix $\mathbf{Z} + \mathbf{Y}_4/\mu$ and Θ is a singular value thresholding operator [17].

4. **Update \mathbf{Z} with fixed other variables.** By dropping the irrelevant terms w.r.t \mathbf{Z} in Eq. 8, we have:

$$\begin{aligned} & \min_{\mathbf{Z}} \langle \mathbf{Y}_3, \mathbf{C} - \mathbf{CZ} - \mathbf{E}_2 \rangle + \langle \mathbf{Y}_4, \mathbf{Z} - \mathbf{J} \rangle + \mu/2 \left(\|\mathbf{C} - \mathbf{CZ} - \mathbf{E}_2\|_F^2 + \|\mathbf{Z} - \mathbf{J}\|_F^2 \right) \\ & = \min_{\mathbf{Z}} \|\mathbf{C} - \mathbf{CZ} - \mathbf{E}_2 + \mathbf{Y}_3/\mu\|_F^2 + \|\mathbf{Z} - \mathbf{J} + \mathbf{Y}_4/\mu\|_F^2 \end{aligned} \quad (14)$$

We also take the derivation of Eq. 14 w.r.t. \mathbf{Z} and set it to $\mathbf{0}$, then the following equation holds:

$$(\mathbf{C}^t\mathbf{C} + \mathbf{I}_n)\mathbf{Z}^{opt} = \mathbf{C}^t(\mathbf{C} - \mathbf{E}_2 + \mathbf{Y}_3/\mu) + \mathbf{J} - \mathbf{Y}_4/\mu. \quad (15)$$

Hence, $\mathbf{Z}^{opt} = (\mathbf{C}^t\mathbf{C} + \mathbf{I}_n)^{-1} \left[\mathbf{C}^t(\mathbf{C} - \mathbf{E}_2 + \mathbf{Y}_3/\mu) + \mathbf{J} - \mathbf{Y}_4/\mu \right]$.

5. **Update \mathbf{E}_1 with fixed other variables.** By abandoning the irrelevant terms of \mathbf{E}_1 , then minimizing Eq. 8 equals solving the following problem:

$$\begin{aligned} & \min_{\mathbf{E}_1} \alpha \|\mathbf{E}_1\|_1 + \langle \mathbf{Y}_1, \mathbf{X} - \mathbf{XC} - \mathbf{E}_1 \rangle + \mu/2 \|\mathbf{X} - \mathbf{XC} - \mathbf{E}_1\|_F^2 \\ & = \min_{\mathbf{E}_1} \alpha \|\mathbf{E}_1\|_1 + \mu/2 \|\mathbf{X} - \mathbf{XC} - \mathbf{E}_1 + \mathbf{Y}_1/\mu\|_F^2. \end{aligned} \quad (16)$$

Similar to computing the optimal value of \mathbf{M} , we could get $[\mathbf{E}_1^{opt}]_{ij} = \max(0, [\mathbf{X} - \mathbf{XC} + \mathbf{Y}_1/\mu]_{ij} - \alpha/\mu) + \min(0, [\mathbf{X} - \mathbf{XC} + \mathbf{Y}_1/\mu]_{ij} + \alpha/\mu)$.

6. **Update \mathbf{E}_2 with fixed other variables.** By gathering the relevant terms of \mathbf{E}_2 in Eq. 8, we have

$$\begin{aligned} & \min_{\mathbf{E}_2} \beta \|\mathbf{E}_2\|_F^2 + \langle \mathbf{Y}_3, \mathbf{C} - \mathbf{CZ} - \mathbf{E}_2 \rangle + \mu/2 \|\mathbf{C} - \mathbf{CZ} - \mathbf{E}_2\|_F^2 \\ & = \min_{\mathbf{E}_2} \beta \|\mathbf{E}_2\|_F^2 + \mu/2 \|\mathbf{C} - \mathbf{CZ} - \mathbf{E}_2 + \mathbf{Y}_3/\mu\|_F^2. \end{aligned} \quad (17)$$

We take the derivation of Eq. 17 w.r.t. \mathbf{E}_2 and set it to $\mathbf{0}$, then the following equation holds:

$$(2\beta + \mu)\mathbf{E}_2^{opt} = \mu(\mathbf{C} - \mathbf{CZ} + \mathbf{Y}_3/\mu). \quad (18)$$

Hence, $\mathbf{E}_2^{opt} = \mu/(2\beta + \mu)(\mathbf{C} - \mathbf{CZ} + \mathbf{Y}_3/\mu)$.

7. **Update parameters with fixed other variables.** The precise updating schemes for parameters existed in Eq. 8 are summarized as follows:

$$\begin{aligned} \mathbf{Y}_1^{opt} &= \mathbf{Y}_1 + \mu(\mathbf{X} - \mathbf{XC} - \mathbf{E}_1), \\ \mathbf{Y}_2^{opt} &= \mathbf{Y}_2 + \mu(\mathbf{C} - \mathbf{M}), \\ \mathbf{Y}_3^{opt} &= \mathbf{Y}_3 + \mu(\mathbf{C} - \mathbf{CZ} - \mathbf{E}_2), \\ \mathbf{Y}_4^{opt} &= \mathbf{Y}_4 + \mu(\mathbf{Z} - \mathbf{J}), \\ \mu^{opt} &= \min(\mu_{max}, \rho\mu), \end{aligned} \quad (19)$$

where μ_{max} and ρ are two given positive parameters.

4.2 Algorithm

The algorithmic procedure of SRR is summarized in Algorithm 1. For a data set, once the solutions to SRR are obtained, SRR defines an affinity graph $[\mathbf{W}]_{ij} = ([\mathbf{Z}]_{ij} + [\mathbf{Z}]_{ji})$, then N-cut is performed on the graph to get segmentation result.

Algorithm 1. Sparse relation representation (SRR)

Input:

Data set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathcal{R}^{D \times n}$, parameters α, β , the maximal number of iteration *Maxiter*;

Output:

The two coefficient matrices $\mathbf{Z}^{opt}, \mathbf{C}^{opt}$, and two noise term $\mathbf{E}_1^{opt}, \mathbf{E}_2^{opt}$;

- 1: Initialize the parameters, i.e., $\mathbf{Y}_1^0 = \mathbf{Y}_2^0 = \mathbf{Y}_3^0 = \mathbf{Y}_4^0 = \mathbf{0}, \mu^0 = 10^{-2}, \mu_{max} = 10^{30}, \rho = 1.1, \varepsilon = 10^{-8}, k = 0$ and $\mathbf{M}^0 = \mathbf{C}^0 = \mathbf{J} = \mathbf{Z}^0 = \mathbf{0}$.
 - 2: **while** $\|\mathbf{X} - \mathbf{X}\mathbf{C}^k - \mathbf{E}_1^k\|_\infty > \varepsilon, \|\mathbf{C}^k - \mathbf{C}^k\mathbf{Z}^k - \mathbf{E}_2^k\|_\infty > \varepsilon$ and $k < \text{Maxiter}$ **do**
 - 3: Update $[\mathbf{M}]_{ij}^{k+1} = \max(0, [\mathbf{C}^k + \mathbf{Y}_2^k/\mu^k]_{ij} - 1/\mu^k) + \min(0, [\mathbf{C}^k + \mathbf{Y}_2^k/\mu^k]_{ij} + 1/\mu^k)$ when $i \neq j$, else $[\mathbf{M}]_{ij}^{k+1} = 0$;
 - 4: Update \mathbf{C}^{k+1} by using Matlab function `lyap()` to solve $(\mathbf{X}^t\mathbf{X} + \mathbf{I}_n)\mathbf{C}^{k+1} + \mathbf{C}^{k+1}(\mathbf{I}_n - \mathbf{Z}^k)(\mathbf{I}_n - (\mathbf{Z}^k)^t) - \mathbf{X}^t(\mathbf{X} - \mathbf{E}_1^k + \mathbf{Y}_1^k/\mu^k) - \mathbf{M}^{k+1} + \mathbf{Y}_2^k/\mu^k - (\mathbf{E}_2^k - \mathbf{Y}_3^k/\mu^k)(\mathbf{I}_n - (\mathbf{Z}^k)^t) = 0$;
 - 5: Update $\mathbf{J}^{k+1} = \mathbf{U}\Theta_{1/\mu^k}(\mathbf{S})\mathbf{V}$, where $\mathbf{U}\mathbf{S}\mathbf{V}$ is the SVD of matrix $\mathbf{Z}^k + \mathbf{Y}_4^k/\mu^k$ and Θ is a singular value thresholding operator;
 - 6: Update $\mathbf{Z}^{k+1} = ((\mathbf{C}^k)^t\mathbf{C}^k + \mathbf{I}_n)^{-1}[(\mathbf{C}^k)^t(\mathbf{C}^k - \mathbf{E}_2^k + \mathbf{Y}_3^k/\mu^k) + \mathbf{J}^{k+1} - \mathbf{Y}_4^k/\mu^k]$;
 - 7: Update $[\mathbf{E}_1^{k+1}]_{ij} = \max(0, [\mathbf{X} - \mathbf{X}\mathbf{C}^{k+1} + \mathbf{Y}_1^k/\mu^k]_{ij} - \alpha/\mu^k) + \min(0, [\mathbf{X} - \mathbf{X}\mathbf{C}^{k+1} + \mathbf{Y}_1^k/\mu^k]_{ij} + \alpha/\mu^k)$;
 - 8: Update $\mathbf{E}_2^{k+1} = \mu^k/(2\beta + \mu^k)(\mathbf{C}^{k+1} - \mathbf{C}^{k+1}\mathbf{Z}^{k+1} + \mathbf{Y}_3^k/\mu^k)$
 - 9: Update $\mathbf{Y}_1^{k+1}, \mathbf{Y}_2^{k+1}, \mathbf{Y}_3^{k+1}, \mathbf{Y}_4^{k+1}, \mu^{k+1}$ by following the updating schedule in Eq. 19;
 - 10: set $k = k + 1$;
 - 11: **end while**
 - 12: **return** the coefficient matrix $\mathbf{Z}^{opt} = \mathbf{Z}^k, \mathbf{C}^{opt} = \mathbf{C}^k, \mathbf{E}_1^{opt} = \mathbf{E}_1^k, \mathbf{E}_2^{opt} = \mathbf{E}_2^k$.
-

4.3 Analyses

Now we discuss the complexity of Algorithm 1. Suppose the data matrix $\mathbf{X} \in \mathcal{R}^{D \times n}$, the complexity of Algorithm 1 is mainly determined by the updating of six variables: $\mathbf{M} \in \mathcal{R}^{n \times n}, \mathbf{C} \in \mathcal{R}^{n \times n}, \mathbf{J} \in \mathcal{R}^{n \times n}, \mathbf{Z} \in \mathcal{R}^{n \times n}, \mathbf{E}_1 \in \mathcal{R}^{D \times n}, \mathbf{E}_2 \in \mathcal{R}^{n \times n}$. We analyze the computational burden of updating these variables in each step.

First, updating \mathbf{M} and \mathbf{E}_1 both need to compute each element of an $n \times n$ matrix, hence their computation burden is $O(n^2)$. Secondly, it takes $O(n^3)$ time to solve a Sylvester equations for updating \mathbf{C} . Third, by performing SVD, the update of \mathbf{J} is $O(n^3)$. Fourthly, updating \mathbf{Z} needs to compute the pseudo-inverse

of an $n \times n$ matrix, whose complexity is $O(n^3)$. Fifthly, it can be easily to find that the computation burden for updating \mathbf{E}_2 is $O(n^2)$. Hence, we can see that the time complexity of Algorithm 1 in each iteration taken together is $O(n^3)$, which is same to that of LRR. Suppose the number of iterations is N , then the total complexity of Algorithm 1 should be $N \times O(n^3)$.

5 Experiments

In this section, subspace segmentation experiments will be performed to verify the effectiveness of SRR. Two types of data sets, such as Hopkins155 motion segmentation database [18], the extended Yale B [19] and ORL face images database [20] will be adopted. The related algorithms including SSC [5], LRR [6], SCLRR¹ [10] are chosen for comparisons.

5.1 Experiments on Hopkins 155 Database

Hopkins155 database [18] is a frequently used benchmark database to test the performances of subspace segmentation algorithms. It consists of 120 sequences of two motions and 35 sequences of three motions. Each sequence is a sole clustering task and there are 155 clustering tasks in total. The features of each sequence were extracted and tracked along with the motion in all frames, and errors were manually removed for each sequence. So it could be regarded that this database only contains slight corruptions. In our experiments, we projected the data to be 12-dimensional by using principal component analysis (PCA) [14]. Figure 2 presents two sample images of Hopkins 155 database.

We performed the four algorithms on Hopkins 155 database and recorded the detailed statistics of the segmentation errors of the four evaluated algorithms including Mean, standard deviation (Std.) and maximal error(Max.) in Table 1. From Table 1, we can see that (1) the mean of segmentation errors obtained by SRR are all slightly better than those of other algorithms; (2) the standard deviation on all data sets obtained by SRR is also superior to those of other algorithms; (3) all the best values are achieved SRR and SCLRR.

5.2 Experiments on Face Image Databases

The brief information of the extended Yale B and ORL face databases are introduced as follows:

The extended Yale B face database contains 38 human faces and around 64 near frontal images under different illuminations per individual. Some images in this database were corrupted by shadow. We just selected images from first 10 classes of the extended Yale B database to form a heavily corrupted subset.

¹ The reasons why we choose SCLRR for comparison are illustrated as follows: firstly, SCLRR is the generalization of NNLSR; secondly, both SCLRR and NNLSR impose the low-rank and sparse constraints on the reconstruction coefficient matrix to hope it could find the local and global structures of data sets.

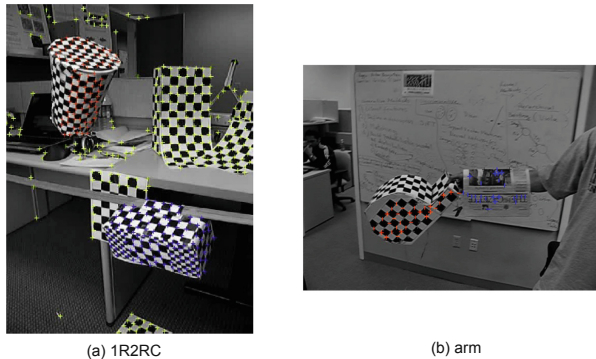


Fig. 2. Sample Images of Hopkins 155 motion segmentation database. (a) 1R2RC, (b) arm.

Table 1. The segmentation errors (%) of different algorithms on Hopkins 155 database. The optimal values of different criterion are emphasized in bold style.

Method	2 motions			3 motions			ALL		
	Mean	Std.	Max.	Mean	Std.	Max.	Mean	Std.	Max.
SSC	4.02	10.24	41.59	11.16	11.00	37.56	5.63	10.81	41.59
LRR	3.13	7.45	30.20	6.56	7.49	23.41	3.93	7.57	30.20
SCLRR	2.96	6.42	31.09	5.18	7.24	24.68	3.49	8.49	29.09
SRR	2.34	6.50	27.23	5.67	7.16	22.41	2.96	5.73	29.23

Notice: The corresponding parameters in different algorithms varied in the interval $[0.01, 10]$. And the best result obtained by each evaluated algorithm on each sub-database was recorded.

ORL database contains 400 face images (without noise) of 40 persons. Each person has 10 different images. These images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). In our experiments, all the images from the extended Yale B and ORL database are resized to 32×32 pixels. Moreover, for effective computation, the element value of each image vector was normalized into the interval $[0,1]$ by being divided 255. Some sample images from the two databases are shown in Fig. 3(a) and (b) respectively.

We performed subspace segmentation experiments on some sub-databases constructed from the above used two image databases. Each sub-database contains the images from q persons (q changes from a relative small number to the total number of class). Then the four algorithms are performed to obtain the subspace segmentation accuracies. In these experiments, all the corresponding parameters in each evaluated algorithm are varied from 0.001 to 20, and the best values corresponding to the highest accuracy of each evaluated algorithm are chosen. Finally, the segmentation accuracy curve of each algorithm against the number of class q are plotted in Fig. 4.

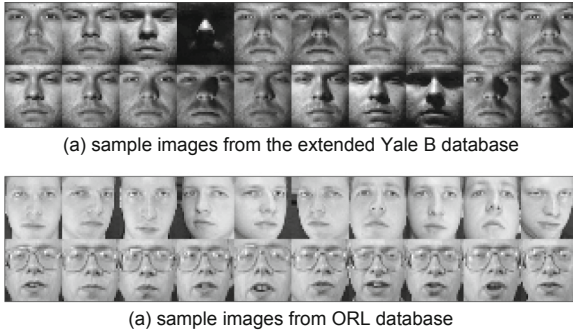


Fig. 3. Sample images from (a) the extended Yale B databases and (b) ORL.

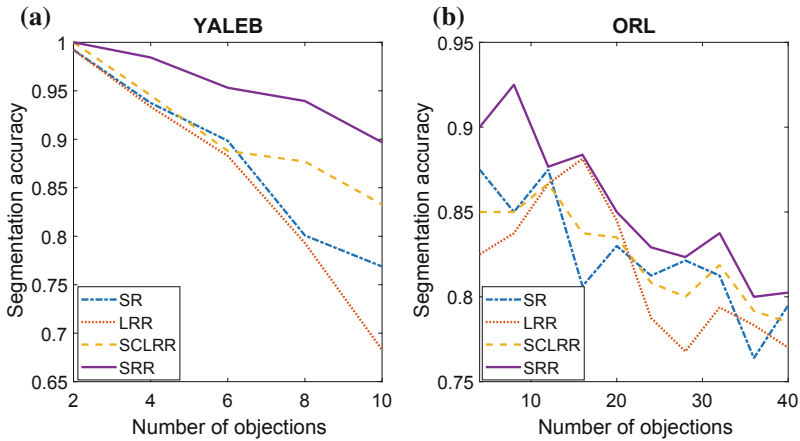


Fig. 4. The segmentation accuracies obtained by the evaluated algorithms versus the variations of number of class on different databases. (a) the extend Yale B (b) ORL databases.

Clearly, from Fig. 4, we can find that (1) in all the experiments, the best results are almost achieved by SRR; (2) the results of SRR are much better than those of other algorithms on the extended Yale B database.

6 Conclusion

We presented the relation representation conception in this paper and developed a kind of sparse relation representation (SRR) method for subspace segmentation. Different from the existing spectral clustering based subspace segmentation algorithms, SRR used the sparse neighborhood relation of each data sample to obtain the affinity graph for a data set. We claimed the obtained affinity graph could discover the subspace structure of the given data set more truthfully. The comparative experiments conducted on several benchmark databases showed that SRR dominated some related algorithms.

References

1. Rao, S., Tron, R., Vidal, R., Ma, Y.: Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(10), 1832–1845 (2010)
2. Ma, Y., Derksen, H., Hong, W., Wright, J.: Segmentation of multivariate mixed data via lossy coding and compression. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(9), 1546–1562 (2007)
3. Wei, L., Wu, A., Yin, J.: Latent space robust subspace segmentation based on low-rank and locality constraints. *Expert Syst. Appl.* **42**, 6598–6608 (2015)
4. Wei, L., Wang, X., Yin, J., Wu, A.: Spectral clustering steered low-rank representation for subspace segmentation. *J. Vis. Commun. Image Represent.* **38**, 386–395 (2016)
5. Elhamifar, E., Vidal, R.: Sparse subspace clustering. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2009, Miami, Florida, USA*, pp. 2790–2797 (2009)
6. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation, In: *Frnkranz, J., Joachims, T. (eds.) Proceedings of the 27th International Conference on Machine Learning, ICML 2010, Haifa, Israel*, pp. 663–670 (2010)
7. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905 (2000)
8. Li, C.-G., Vidal, R.: Structured sparse subspace clustering: a unified optimization framework. In: *CVPR* (2015)
9. Chen, H., Wang, W., Feng, X.: Structured sparse subspace clustering with within-cluster grouping. *Pattern Recognit.* **83**, 107–118 (2018)
10. Zhuang, L., Gao, H., Lin, Z., Ma, Y., Zhang, X., Yu, N.: Non-negative low rank and sparse graph for semi-supervised learning. In: *CVPR*, pp. 2328–2335 (2012)
11. Tang, K., Liu, R., Zhang, J.: Structure-constrained low-rank representation. *IEEE Trans. Neural Netw. Learn. Syst.* **25**, 2167–2179 (2014)
12. Lu, X., Wang, Y., Yuan, Y.: Graph-regularized low-rank representation for destriping of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **51**(7–1), 4009–4018 (2013)
13. Lin, Z., Chen, M., Wu, L., Ma, Y.: The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, UIUC, Champaign, IL, USA, Technical report UILU-ENG-09-2215 (2009)
14. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley, Hoboken (2001)
15. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323–2326 (2000)
16. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009)
17. Cai, J.F., Candes, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **20**(4), 1956–1982 (2010)
18. Tron, R., Vidal, R.: A benchmark for the comparison of 3D motion segmentation algorithms. In: *CVPR* (2007)
19. Lee, K., Ho, J., Driegman, D.: Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(5), 684–698 (2005)
20. Samaria, F., Harter, A.: Parameterisation of a stochastic model for human face identification (1994)