International Journal of Grid Distribution Computing
Vol.8, No.1 (2015), pp.259-268
http://dx.doi.org/10.14257/ijgdc.2015.8.1.24

Study on Different Representation Methods for Subspace Segmentation

Jiangshu Wei, Mantao Wang and Qianqian Wu
College of Information and Engineering, Sichuan Agricultural University, Ya'an, 625014, China
sicauwjs@163.com, mantaowang@163.com, wuqianqian266@163.com

Abstract

With many engineering and science application problems, we must deal with a lot of high-dimensional data, such as videos, images, web documents, text, etc. In the areas of computer vision, image processing and machine learning, high-dimensional data are widespread. However, it is very hard for obtaining meaningful learning and inference from these high-dimensional data directly, the computational complexity of high-dimensional data is often exponential. However, under many conditions, high-dimensional data lie in low-dimensional data corresponding to some classes of the data. Thus, finding the low-dimensional structure from the high-dimensional data is very important. The aim of subspace segmentation is to cluster data that lie in a union of low-dimensional subspaces. In recent years, based on the research of representation methods, many subspace segmentation algorithms appeared. Although these methods are all effective for handling subspace segmentation problems, they all have advantages and disadvantages. This paper focuses on the performance comparison of different subspace segmentation algorithms currently used in handling subspace segmentation problems and views other conventional methods that can be applied in this field.

Keywords: *High-dimensional, Subspace segmentation, Low-dimensional, Representation methods*

1. Introduction

With the rapid development of information technology and computer network, many application problems of computer vision, image processing and machine learning have been studied in many literatures [1-17]. However, many data in these areas are high-dimensional, such as images, videos, web documents, *etc*. It is a little hard for obtaining meaningful inference from these high-dimensional data. As a matter of fact, high-dimensional data often lie in low-dimensional data corresponding to some categories of the data. The aim of subspace segmentation methods is to cluster data that lie in a union of low-dimensional subspaces. Thus, the research of subspace segmentation methods is a hot topic in recent years [1, 2].

Given a collection of data points from some classes lying in a union of unknown subspaces, the aim of subspace segmentation task is to separate the data points according to the underlying subspaces. Subspace segmentation has been widely used in scientific and engineering fields, such as pattern recognition, machine learning, image processing, computer vision, *etc*. Figure 1 is a classical application of subspace segmentation [1, 2].

Figure 1. Given Face Images from Different Persons, the Aim of Subspace Segmentation is to Find the Images that Belong to the Same Person

ISSN: 2005-4262 IJGDC

There are also many conventional methods for handling subspace segmentation problems, such as Generalized Principal Component Analysis (GPCA), Random Sample Consensus (RANSAC), Local Subspace Affinity (LSA), *etc* [1, 2]. Recently, there has been an increasing interest in representation theory. Representation methods include many algorithms, such as sparse representation, low rank representation, *etc*. Based on the research of representation methods, many subspace segmentation algorithms have appeared. Although these methods have their own characteristics, they are all effective for handling subspace segmentation problems.

However, subspace segmentation methods also include many algorithms. These algorithms also have advantages and disadvantages. The aim of this paper is to compare the state of the art algorithms in subspace segmentation. We try to clarify the similarities among different subspace segmentation algorithms and reveal the differences of them.

Based on the research of sparse representation, Elhamifar and Vidal proposed the Sparse Subspace Clustering (SSC) method [1, 2]. The solution of SSC is obtained by using l_1 norm minimization. SSC boosts the research of subspace segmentation. Many application problems are solved by the SSC method.

Based on SSC, some papers proposed other methods for improving it. Soltanolkotabi *et al.* [3] proposed an extensive algorithm of SSC. They proposed a two-step procedure with data-driven regularization method, which is a extensive method of SSC. This method is a statistical mixture model to represent data points lying in a union of subspaces. The performance of the method is also strictly explained according to interpretable and intuitive parameters.

Although SSC is effective for handling some subspace segmentation applications, it also has some problems. SSC is based on the sparse representation, sparse representation only finds the sparsest representation of each data vector, thus, there is no global constraint on its solution. So when the data is grossly corrupted, SSC is inaccurate for capturing the global structures of the data points. In view of the disadvantages of the SSC, some other novel methods are proposed. The low rank representation (LRR) was proposed by Liu [4], which is different from the sparse representation. The aim of LRR is to find the lowest-rank representation of all data points jointly. Thus, LRR can be better at capturing the global structures of data points. For corrupted data, the lowest-rank representation can correct corruptions so that LRR is robust to noise. Based on LRR, some other improvement methods were also proposed.

This paper is organized as follows. In section 2, SSC method is reviewed. In section 3, the extensive method of SSC is reviewed. In section 4, low rank representation method is reviewed. In section 5, based on LRR, some improvement methods are reviewed. In section 6, some experiments have been done in order to compare these subspace segmentation methods. In section 7, we review the advantages and disadvantages of these methods. Section 8 concludes this paper.

2. Sparse Subspace Clustering (SSC) Method

2.1. Problem Formulation

Given a set of data points lying in a union of unknown linear subspaces, there are n subspaces $S_1, S_2, ..., S_n$ of R^D of dimensions $d_1, d_2, ..., d_n$. Assume we are given a set of N data points $y_1, y_2, ..., y_N$, the aim of subspace segmentation is to approximate the underlying subspaces by using the set of data points [1], [2], [3].

2.2. The Steps of SSC Method

SSC is based on the theory of sparse representation, which takes into account of the self-expressions property of the data points. Under some conditions**,** SSC method is

effective for handling subspace segmentation problems. The steps of SSC are as follows [1], [2]:

Input: Assume we are given a set of N data points $y_1, y_2, ..., y_N$ lying near a union of n subspaces $S_1, S_2, ..., S_n$ of R^D of dimensions $d_1, d_2, ..., d_n$.

1): Solve the sparse representation optimization problem:

$$\min \|B\|_1 \quad \text{s.t. } Y = YB,\ diag(B) = 0.$$

2): Construct a similarity graph G with nodes representing the *N* data points, the node *i* is represented y_i, the node *j* is represented y_j. The edge weight of G is given by $w_{ij} = |b_{ij}| + |b_{ji}|$.

3): Sort the eigenvalues of the normalized Laplacian matrix of the graph G in descending order.

4): Apply the spectral clustering to the similarity graph in order to obtain the partition $Y_1, Y_2, ..., Y_n$.

Output: Obtain the segmentation of the data points: $Y_1, Y_2, ..., Y_n$.

3. The Extensive Method of SSC

SSC is a classical method for dealing with the subspace segmentation problems, however, there are also some disadvantages about SSC. Thus, some improvement methods of SSC were also proposed. Soltanolkotabi *et al.* [3] proposed an extensive algorithm of SSC. They proposed a two-step procedure with data-driven regularization method, which is an extensive method of SSC. The performance of the method is also strictly explained according to interpretable and intuitive parameters, such as the affinity between subspaces, sampling density and the noise level, *etc*. The steps of two-step procedure with data-driven regularization method are as follows:

for *i*=1,…,*N* do

1): Solve $\alpha^* = \arg\min_{\alpha \in R^D} \|\alpha\|_{l_1}$ s.t. $\|y_i - Y\alpha\|_{l_2} \le \tau$ and $\alpha_i = 0$.

2): Set $\lambda = f(\|\alpha^*\|_{l_1})$.

3): Solve $\hat{\alpha} = \arg\min_{\alpha \in R^D} \frac{1}{2}\|y_i - Y\alpha\|_{l_2}^2 + \lambda\|\alpha\|_{l_1}$ s.t. $\alpha_i = 0$.

4): Set $B_i = \hat{\alpha}$.

end for

4. Low Rank Representation (LRR) methods

4.1. Problem Formulation

SSC and its extensive method are all based on the theory of sparse representation, the sparse representation only looks for the sparsest representation of each data vector individually. However, the sparsest solution does not mean obtaining the highest accuracy at capturing the global structures of the data points [4]. Thus, when the data points are grossly corrupted, the performance of SSC will be degraded. In view of the disadvantages of SSC and its extensive methods, some other novel methods were proposed.

Liu *et al*. [4] proposed the Low Rank Representation (LRR) method, which is different from sparse representation. The aim of the sparse representation is to obtain the sparsest

representation of each data vector respectively. However, unlike sparse representation, the aim of low rank representation is to find the lowest rank representation of all data points jointly. Compared to SSC, LRR can be better at capturing the global structures of data points. The lowest-rank criterion can correct corruptions so that LRR is robust to noise and outliers.

With a set of clean data vectors $Y = [y_1, y_2, ..., y_N]$ drawn from a union of n subspaces $S_1, S_2, ..., S_n$, every column is a data vector. Let $d_1, d_2, ..., d_n$ be the unknown dimensions of the *n* subspaces. Y_i be the collection of n_i data vectors which are from the *i*-th subspace S_i . Our aim is to capture the global structure of Y . However, sparse representation cannot capture the global structure of Y , low rank representation is a more appropriate criterion for capturing the global structure of Y . Here, the data set Y is used as the dictionary, the LRR problem is to solve the following optimization problem:

$$\min_{Z} \|Z\|_* \quad \text{s.t.} \quad Y = YZ . \qquad (1)$$

For problem (1), $\|\bullet\|_*$ is the nuclear norm of a matrix, the definition of the nuclear norm is the sum of the singular values of the matrix.

4.2. Under the Conditions that the Data Vectors are Noisy

In many real applications, the data vectors are often noisy. Assume that a fraction of the data vectors are corrupted, it means the data vectors $Y = [y_1, y_2, ..., y_N]$ are corrupted, $Y = D + E$, *Y* are the corrupted data vectors, *D* are the clean data vectors and *E* are the noise vectors. Under this condition, the optimization objective function is as follows [4]:

$$\min_{Z,E} \|Z\|_* + \lambda \|E\|_{2,1}, \ \text{s.t.} \ Y = YZ + E , \quad (2)$$

$\|E\|_{2,1} = \sum_{j=1}^{N} \sqrt{\sum_{i=1}^{N} (E_{ij})^2}$ is the $l_{2,1}$ norm, the parameter $\lambda > 0$ is used to balance the effects of the two parts, which could be tuned empirically. This method of (2) was called LRR_1.

Sometimes, $\|E\|_{2,1}$ can be substituted by $\|E\|_1$, the optimization objective function is changed as follows:

$$\min_{Z,E} \|Z\|_* + \lambda \|E\|_1, \ \text{s.t.} \ Y = YZ + E , \quad (3)$$

this method of (3) was called $LRR_{2,1}$.

The problem (1), (2) and (3) can be solved by using Augmented Lagrange Multiplier (ALM) algorithm [18-22], which is a classical method for solving the Low Rank Representation (LRR) problem. However, there are also some disadvantages with LRR. First, it cannot obtain closed form solution from the ALM algorithm. Second, there are too many parameters with the ALM algorithm. Third, the convergence property of ALM cannot be analyzed in detail.

5. Other Improvement Methods based on LRR

In view of the advantages and disadvantages of LRR, based on the theory of low rank representation, some other improvement methods were also proposed.

5.1 Robust Shape Interaction (RSI) Method

Wei and Lin proposed the Robust Shape Interaction (RSI) method [5], which is an improved method of LRR. The most important thing of RSI is that it uses the corrected data as the dictionary instead of the noisy data. Thus, RSI is more robust than LRR when the data vectors are grossly corrupted. LRR uses the noisy data set Y itself as the dictionary instead of the clean data D. Wei and Lin considered that using the noisy data set as the dictionary was not quite reasonable when the noise was heavy. If using the clean data as the dictionary, the problem (2) should be changed as follows:

$$\min_{Z,D,E} \|Z\|_* + \lambda \|E\|_{2,1}, \quad \text{s.t. } D = DZ,\ Y = D + E. \qquad (4)$$

From one theorem proposed by Wei and Lin, (4) is equivalent to

$$\min_{D,E} rank(D) + \lambda \|E\|_{2,1}, \quad \text{s.t.} \quad Y = D + E. \qquad (5)$$

Replace the rank function in (5) with the nuclear norm, the core of RSI method is to solve the following convex optimization problem:

$$\min_{D,E} \|D\|_* + \lambda \|E\|_{2,1}, \quad \text{s.t.} \quad Y = D + E, \qquad (6)$$

the problem (6) can be also solved by using Augmented Lagrange Multiplier (ALM) algorithm.

5.2 Least Squares Regression (LSR) Method

The LSR method was proposed by Lu *et al* [6]. They considered that many existing methods were all special cases, then they presented the Least Squares Regression (LSR) method for subspace segmentation. The core of LSR method is that it takes advantage of data correlation,
which is common in real data. They believed that most data exhibited strong correlations. If the sampling data were sufficient, they tended to be highly correlated. The Least Squares Regression (LSR) method is as follows:

$$\min \|Z\|_F \quad \text{s.t.} \quad Y=YZ, \quad \text{diag}(Z)=0, \qquad (7)$$

where $\|Z\|_F$ denotes the Frobenius norm of Z, $\|Z\|_F = (\sum_{i=1}^{N}\sum_{j=1}^{N} Z_{ij}^2)^{\frac{1}{2}}$.

In many real applications, the data vectors are often noisy, under this condition, the LSR method is changed as follows:

$$\min \|Y - YZ\|_F^2 + \lambda \|Z\|_F^2 \ \text{s.t. diag}(Z)=0, \qquad (8)$$

where $\lambda > 0$ is used to balance the effects of the two parts, from the optimization problem (8), a closed form solution can be obtained, which gives $Z^* = -D(diag(D))^{-1}$ and $diag(Z^*) = 0$, where $D = (Y^T Y + \lambda I)^{-1}$.

5.3 Other Closed form Solution Method based on LRR

Based on LRR, Favaro *et al*. proposed a method for subspace clustering and segmentation in the presence of noise [7]. The optimization function is as follows:

$$\min_{Z,D,E} \|Z\|_* + \frac{\alpha}{2}\|E\|_F^2, \quad \text{s.t. } D = DZ,\ Y = D + E. \qquad (9)$$

From (9), a closed form solution can be also obtained.

6. Experiments

In this section, some experiments are presented to show the accuracy of subspace segmentation. We focus on the performance comparison of different representation

methods have been mentioned above.

In the experiment, we demonstrate an example of using LRR to do subspace segmentation. We construct 5 independent subspaces, each of which has a rank of 100. 200 points of dimension 100 are sampled from each subspace, and randomly choose some points to corrupt.

In this experiment, we focus on the performance comparison of different subspace segmentation methods. Table 1 and Figure 2 show the segmentation accuracy versus different λ by LRR_1, Table 2 and Figure 3 show the segmentation accuracy versus different λ by $LRR_{2,1}$.

Table 1.The Segmentation Accuracy versus Different λ by LRR_1

λ	0.0001	0.0002	0.003	0.01	0.03	0.06	0.1
accuracy	96.9%	96.9%	80%	78%	78%	64.6%	77.2%

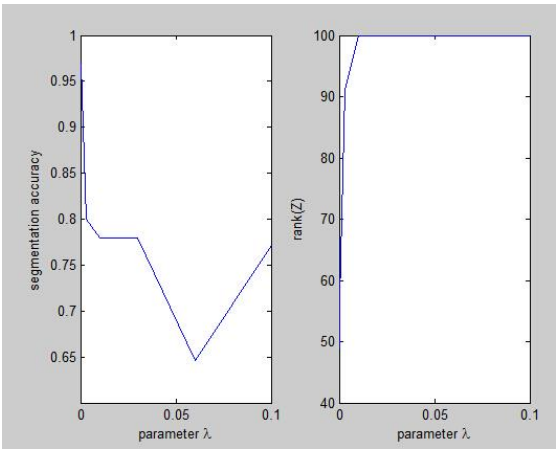

Figure 2. The Segmentation Accuracy and Rank of Solution versus Different λ by LRR1

Table 2. The Segmentation Accuracy versus Different λ by LRR2,1

λ	0.0001	0.0002	0.003	0.01	0.03	0.06	0.1
accuracy	95.5%	95.7%	96.1%	95.9%	76.4%	74.7%	74.7%

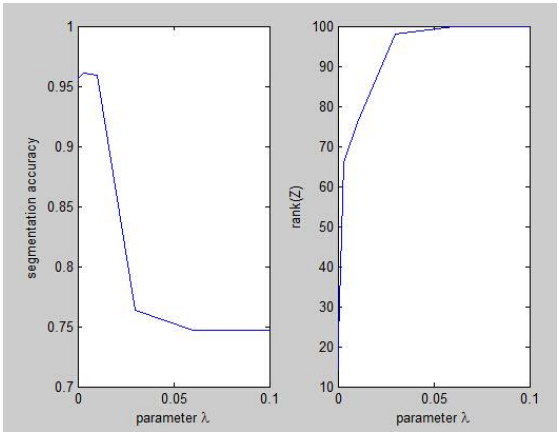

Figure 3. The Segmentation Accuracy and Rank of Solution versus Different λ by LRR2,1

7. The Comparison of Different Methods

As mentioned above, there are many representation methods used for handling the subspace segmentation problems. However, these methods all have their own characteristics, they also have advantages and disadvantages. The reviews of these representation methods are listed in Table 3.

Table 3.The Reviews of Different Representation Methods used for Subspace Segmentation

Method	Characteristics
SSC	Based on the theory of sparse representation, inaccurate at capturing the global structures of data vectors, not be robust to noise and outliers.
Extensive SSC	An extensive method of SSC. The performance of the method is also strictly explained according to interpretable and intuitive parameters.
LRR	Can capture the global structure of samples, cannot obtain closed form solution, too many parameters, the convergence property of ALM cannot be analyzed in detail.
RSI	An improved method of LRR. The most important thing of RSI is that it uses the corrected data as the dictionary instead of the noisy data. RSI is more robust than LRR when the data vectors are grossly corrupted.
LSR	It takes advantage of data correlation, can obtain closed form solution.

8. Conclusions

Data mining technologies are increasingly becoming a hot research field in recent years. However, data mining technologies include many aspects, subspace segmentation is one of the most important aspects of data mining. Representation methods are very effective for handling subspace segmentation problems. However, these methods all have their own characteristics; they also have advantages and disadvantages. In the future work, we should combine the advantages of currently used methods. It means that, the improved representation methods, which the segmentation accuracy should be high and the computational complexity should be low. These methods are expected to apply in subspace segmentation area.

Acknowledgements

This work was supported by the key fund project of Sichuan Educational department (NO.14ZA0005).

References

[1] E. Elhamifar and R.Vidal, "Sparse subspace clustering, Computer Vision and Pattern Recognition (CVPR), IEEE Conference, IEEE, **(2009)**.
[2] E. Elhamifar and R.Vidal, "Sparse subspace clustering: Algorithm, theory, and applications", Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 35, IEEE, no. 11, **(2013)**, pp. 2765-2781.
[3] M. Soltanolkotabi, E. Elhamifar and E. J. Candes, "Robust subspace clustering", The Annals of Statistics, vol. 42, no. 2, **(2014)**, pp.669-699.
[4] G. Liu, Z. Lin and Y. Yu, "Robust subspace segmentation by low-rank representation", Proceedings of the 27th International Conference on Machine Learning (ICML-10), **(2010)**.
[5] S. Wei and Z. Lin, "Analysis and Improvement of Low Rank Representation for Subspace segmentation, arXiv preprint arXiv:1107.1561, **(2011)**.
[6] C. Y. Lu, H. Min, Z. Q. Zhao, L. Zhu, D. S. Huang and S. Yan, "Robust and Efficient Subspace Segmentation via Least Squares Regression", European Conference on Computer Vision (ECCV), **(2012)**.
[7] P. Favaro, R. Vidal and A. Ravichandran, "A closed form solution to robust subspace estimation and clustering", Computer Vision and Pattern Recognition (CVPR), IEEE Conference, IEEE, **(2011)**.
[8] J. J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise", Information Theory, IEEE Transactions, vol. 51, no. 10, **(2005)**, pp. 3601–3608.
[9] T. Guha and R. K. Ward, "Learning sparse representations for human action recognition", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 8, **(2012)**, pp. 1576–1588.
[10] J. Mairal, F. Bach, J. Ponce and G. Sapiro, "Online dictionary learning for sparse coding", Proceedings of the 26th Annual International Conference on Machine Learning, **(2009)**.
[11] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", Science, vol. 290, no. 5500, **(2000)**, pp. 2323–2326.
[12] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifoldsV", The Journal of Machine Learning Research, vol. 4, **(2003)**, pp. 119–155.
[13] M. Stojnic, F. Parvaresh and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements", Signal Processing, IEEE Transactions, vol. 57, no. 8, **(2009)**, pp. 3075–3085.
[14] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation", Information Theory, IEEE Transactions, vol. 50, no. 10, **(2004)**, pp. 2231–2242.
[15] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang and S. Yan, "Sparse representation for computer vision and pattern recognition", Proceedings of the IEEE, **(2010)**.
[16] M. Soltanolkotabi, E. Elhamifar and E. J. Candes, "Robust Subspace Clustering", arXiv preprint arXiv:1301.2603, **(2013)**.
[17] K. Lee, J. Ho and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, no. 5, **(2005)**, pp. 684–698.
[18] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction", Computer Vision (ICCV), IEEE International Conference, **(2011)**.
[19] Z. Shi, J. Han, T. Zheng and S. Deng "Online Learning for Classification of Low-rank Representation Features and Its Applications in Audio Segment Classification", arXiv preprint arXiv:1112.4243, **(2011)**.
[20] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu and Y. Ma, "Robust recovery of subspace structures by low-rank representation", Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 35, no. 1, **(2013)**, pp. 171-184.
[21] G. Ye, D. Liu, I. Jhuo and S. Chang, "Robust late fusion with rank minimization, Computer Vision and Pattern Recognition (CVPR), IEEE Conference, **(2012)**.
[22] Z. Lin, M. Chen and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices", arXiv preprint arXiv:1009.5055, **(2010)**.

Authors

Jiangshu Wei. He received his B.S. degree in Sichuan Normal University in 2003. He received his M.S. degree in University of Electronic Science and Technology of China in 2006. Now, he is a lecturer at the College of information engineering, Sichuan Agricultural University. He is also a Ph.D candidate in the college of computer science at Sichuan University. His main research interests include data mining, pattern recognition and machine learning.

Mantao Wang, he received the B.S. degree in Computer Science and Technology from University of Electronic Science and Technology of China in 1997. He is a lecturer at the School of information engineering, Sichuan Agricultural University. His research interests include image processing and pattern recognition.

Qianqian Wu, she received the B.E. degree in Computer Science and Technology from Sichuan University in 2003. She is a lecturer at the School of information engineering, Sichuan Agricultural Univers ity. Her research interests include image processing and data mining.