Fast Spectral Clustering With Anchor Graph for Large Hyperspectral Images

Rong Wang, *Member, IEEE*, Feiping Nie, and Weizhong Yu

***Abstract*— The large-scale hyperspectral image (HSI) clustering problem has attracted significant attention in the field of remote sensing. Most traditional graph-based clustering methods still face challenges in the successful application of the large-scale HSI clustering problem mainly due to their high computational complexity. In this letter, we propose a novel approach, called fast spectral clustering with anchor graph (FSCAG), to efficiently deal with the large-scale HSI clustering problem. Specifically, we consider the spectral and spatial properties of HSI in the anchor graph construction. The proposed FSCAG algorithm first constructs anchor graph and then performs spectral analysis on the graph. With this, the computational complexity can be reduced to $O(ndm)$, which is a significant improvement compared to conventional graph-based clustering methods that need at least $O(n^2d)$, where n, d, and m are the number of samples, features, and anchors, respectively. Several experiments are conducted to demonstrate the efficiency and effectiveness of the proposed FSCAG algorithm.**

***Index Terms*— Anchor graph, graph-based clustering, hyperspectral image (HSI).**

I. Introduction

HYPERSPECTRAL images (HSIs) are acquired by imaging and spectroscopy technology, and these images provide both rich spectral and spatial information for monitoring the earth's surface [1], [2]. Therefore, HSI has been widely applied in many areas including precision agriculture, military applications, and environmental management. Among these applications, HSI clustering is extremely important and has attracted significant attention in recent years. The aim of HSI clustering is to partition a given image into groups such that pixels in the same group are as similar to each other as possible, while those assigned to different groups are dissimilar. However, HSI clustering is a very challenging task for the large spectral variability and complex spatial structures of HSI.

So far, many clustering methods with various working mechanisms for HSI have been proposed. Zhang *et al.* [3] summarize the existing clustering algorithms for HSI and divide them into the following categories: 1) centroid-based methods, such as k-means [4], fuzzy c-means (FCM) [5], and FCM_S1 [6]; 2) density-based methods, such as clustering by fast search and find of density peaks [7] and the clustering-in-quest method [8]; 3) biological clustering methods, such as remote sensing unsupervised artificial immune networks [9] and the automatic fuzzy clustering method based on adaptive multiobjective differential evolution [10]; and 4) graph-based methods, such as spectral clustering (SC) [11], spectral curvature clustering [12], and sparse subspace clustering [3], [13].

In this letter, we focus on the family of graph-based clustering methods. Most traditional graph-based clustering methods, however, focus on clustering accuracy while ignoring the underlying computational complexity, which is of great importance for the clustering of large-scale HSI. The complexity mainly arises from two aspects. The first is the similarity matrix construction, which takes $O(n^2d)$ time complexity, and the second is the eigenvalue decomposition of the Laplacian matrix, which is implemented in $O(n^2c)$, where n, d, and c are the number of samples, features, and classes, respectively. Both of them are an unbearable burden for the large-scale HSI clustering problem.

To alleviate the above problem, inspired by recent work in scaling up graph-based learning models by anchors [14]–[18], we propose a novel approach, called fast SC with anchor graph (FSCAG), to efficiently deal with the large-scale HSI clustering problem. The main contributions of this letter are as follows. First, the anchor graph is constructed by considering the high spectral correlation and rich spatial information of the HSI. Second, the overall computational complexity of the FSCAG algorithm is $O(ndm)$, which has great advantage over conventional graph-based clustering methods. Comprehensive experiments on several large-scale HSI data sets demonstrate the efficiency and effectiveness of the FSCAG algorithm in both the quantitative and visual evaluations.

II. FSCAG Algorithm

In this section, we introduce our FSCAG for large-scale HSI clustering. The anchor graph is first constructed by incorporating the spatial information of HSI. Then, we design an efficient way for spectral analysis on this graph.

A. Anchor Graph Construction

Let $\boldsymbol{X} = [x_1, \ldots, x_n]^T \in \mathbb{R}^{n\times d}$ denote the data matrix, where n is the number of data points, d is the dimension of features, and each data point $x_i \in \mathbb{R}^d$ belongs to one

Manuscript received April 22, 2017; revised July 13, 2017 and August 1, 2017; accepted August 25, 2017. Date of publication September 18, 2017; date of current version October 25, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61401471 and Grant 61772427 and in part by the China Post-Doctoral Science Foundation under Grant 2014M562636. *(Corresponding author: Feiping Nie.)*

The authors are with the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: wangrong07@tsinghua.org.cn; feipingnie@gmail.com).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2017.2746625

1545-598X © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

of c classes. Given a data matrix $\boldsymbol{X}$, each data point x_i is represented as a vertex on the affinity graph and each edge represents the similarity relationship of one pair of vertexes. The weight of the edge between x_i and x_j is defined as a_{ij}, and $\boldsymbol{A} = \{a_{ij}\} \in \mathbb{R}^{n\times n}, \forall i, j \in 1, \ldots, n$ denotes the similarity matrix of the affinity graph.

Recent studies [14]–[18] adopt an anchor-based strategy to construct similarity matrix $\boldsymbol{A}$. In general, the anchor-based strategy first generates m anchors from data points, where $m \ll n$, and then designs a matrix $\boldsymbol{Z} \in \mathbb{R}^{n\times m}$ that measures the similarity between data points and anchors.

The first step of the anchor-based strategy is anchors generation, which can be achieved by random selection or using the k-means method. Random selection selects m anchors by random sampling from data points with computational complexity $O(1)$. Although random selection cannot guarantee that the selected m anchors are always good, it is extremely fast for large-scale HSI clustering. The k-means method makes use of m clustering centers as anchors. The clustering centers are more representative anchors, but the computational complexity of k-means is $O(ndmt)$, where t is the number of iterations, which makes it impossible for large-scale HSI clustering.

Let $\boldsymbol{U} = [u_1, \ldots, u_m]^T \in \mathbb{R}^{m\times d}$ denote the generated anchors. The (i, j)th element of $\boldsymbol{Z}$ is defined as [14]

$$z_{ij} = \frac{K(x_i, u_j)}{\sum_{s\in\Phi_i} K(x_i, u_s)}, \quad \forall j \in \Phi_i \tag{1}$$

where $\Phi_i \subset \{1, \ldots, m\}$ denotes the indexes of k nearest neighbors of x_i in $\boldsymbol{U}$, $K()$ is a given kernel function, and Gaussian kernel $K(x_i, u_j) = \exp(-\|x_i - u_j\|_2^2/2\sigma^2)$ is adopted in [14], but kernel-based methods always bring extra parameters, e.g., bandwidth σ. To avoid this, Nie *et al.* [18] obtain the ith row of $\boldsymbol{Z}$ by solving the following problem:

$$\min_{z_i^T\mathbf{1}=1, z_{ij}\geq 0} \sum_{j=1}^{m} \|x_i - u_j\|_2^2 z_{ij} + \gamma z_{ij}^2 \tag{2}$$

where z_i^T denotes the ith row of $\boldsymbol{Z}$ and γ is a regularization parameter. Equation (2) is parameter-free, which has been proven in [19], but it does not take the spatial information of HSI into account, which may lead to the result that some isolated pixels appear in the clustering HSI image due to the existence of noise, outliers, or mixed pixels [10]. Inspired by the work in [6] and [20], we propose to incorporate the spatial information by directly modifying the objective function in (2) as follows:

$$\min_{z_i^T\mathbf{1}=1, z_{ij}\geq 0} \sum_{j=1}^{m} \|x_i - u_j\|_2^2 z_{ij} + \alpha\|\bar{x}_i - u_j\|_2^2 z_{ij} + \gamma z_{ij}^2 \tag{3}$$

where $\bar{x}_i$ is the mean of the neighboring pixels lying within a window around x_i and can be calculated in advance. The parameter α controls the tradeoff between the original HSI and the corresponding mean-filtered HSI. Denote $d_{ij}^x = \|x_i - u_j\|_2^2$ and $d_{ij}^{\bar{x}} = \|\bar{x}_i - u_j\|_2^2$, and denote $d_i \in \mathbb{R}^{m\times 1}$ is a vector with the jth element as $d_{ij} = d_{ij}^x + \alpha d_{ij}^{\bar{x}}$. Then problem (3) can be written in vector form as

$$\min_{z_i} \left\|z_i + \frac{1}{2\gamma} d_i\right\|_2^2 \quad \text{s.t. } z_i^T\mathbf{1} = 1,\ z_{ij} \geq 0. \tag{4}$$

Following the work in [19], it is preferred to learn a sparse z_i that has exactly k nonzero values. Thus, the learned $\boldsymbol{Z}$ is sparse, and the computation burden of subsequent spectral analysis can be alleviated largely. The parameter γ can be set as $\gamma = (k/2)d_{i,k+1} - (1/2)\sum_{j=1}^{k} d_{i,j}$, and the solution to (4) is

$$z_{ij} = \frac{d_{i,k+1} - d_{ij}}{k d_{i,k+1} - \sum_{j=1}^{k} d_{ij}}. \tag{5}$$

For detail derivation, the reader is referred to [19]. Thus, we only need $O(ndm)$ to compute the matrix $\boldsymbol{Z}$. Then, the similarity matrix $\boldsymbol{A}$ can be obtained by [14]

$$\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{\Lambda}^{-1}\boldsymbol{Z}^T \tag{6}$$

where the diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times m}$ is defined as $\boldsymbol{\Lambda}_{jj} = \Sigma_{i=1}^{n} z_{ij}$.

B. Spectral Analysis on Anchor Graph

The objective function of SC is

$$\min_{\boldsymbol{F}^T\boldsymbol{F}=\boldsymbol{I}} \operatorname{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}) \tag{7}$$

where $\boldsymbol{F} \in \mathbb{B}^{n\times c}$ is the clustering indicator matrix and c is the number of classes. $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ is called a Laplacian matrix in graph theory, and the degree matrix $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ is defined as a diagonal matrix where the ith diagonal element is $\sum_{j=1}^{n} a_{ij}$. Note that the elements of $\boldsymbol{F}$ are constrained to be discrete values, which makes (7) hard to solve. A well-known solution to this problem is to relax the matrix $\boldsymbol{F}$ from the discrete values to continuous ones. By performing eigenvalue decomposition on $\boldsymbol{L}$, we obtain the relaxed continuous solution that is composed of eigenvectors corresponding to the smallest c eigenvalues. Then, k-means clustering can be used to calculate the discrete solution.

However, the computational complexity of eigenvalue decomposition on $\boldsymbol{L}$ is $O(n^2c)$, which is not suitable for large-scale HSI clustering. From (6), the (i, j)th element of $\boldsymbol{A}$ is $a_{ij} = z_i^T\boldsymbol{\Lambda}^{-1}z_j$, which satisfies $a_{ij}^T = a_{ji}$. Thus $\boldsymbol{A}$ is a symmetric matrix. Note that $z_{ij} \geq 0$ and $z_i^T\mathbf{1} = 1$, and it can be verified that $\boldsymbol{A}$ is a double stochastic matrix [21] that satisfies $a_{ij} \geq 0$ and

$$\sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} z_i^T\boldsymbol{\Lambda}^{-1}z_j = z_i^T\sum_{j=1}^{n}\boldsymbol{\Lambda}^{-1}z_j = z_i^T\mathbf{1} = 1. \tag{8}$$

Thus, the similarity matrix $\boldsymbol{A}$ is automatically normalized and the degree matrix $\boldsymbol{D} = \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the identity matrix. Since $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{A}$, problem (7) is equivalent to the following problem:

$$\max_{\boldsymbol{F}^T\boldsymbol{F}=\boldsymbol{I}} \operatorname{tr}(\boldsymbol{F}^T\boldsymbol{A}\boldsymbol{F}). \tag{9}$$

According to (6), $\boldsymbol{A}$ can be written as $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}^T$, where $\boldsymbol{B} = \boldsymbol{Z}\boldsymbol{\Lambda}^{-(1/2)}$. The singular value decomposition (SVD) of $\boldsymbol{B}$ can be written as follows:

$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{10}$$

where $\boldsymbol{U} \in \mathbb{R}^{n\times n}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times m}$, and $\boldsymbol{V} \in \mathbb{R}^{m\times m}$ are the left singular vector matrix, singular value matrix,

and right singular vector matrix, respectively. It is easy to verify that the column vectors of $\boldsymbol{U}$ are the eigenvectors of the similarity matrix $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}^T$. Instead of directly performing eigenvalue decomposition on $\boldsymbol{A}$, we prefer to perform SVD on $\boldsymbol{B}$ to obtain the relaxed continuous solution of $\boldsymbol{F}$, and the computational complexity reduces to $O(nmc + m^2c)$. After that, k-means clustering is used to calculate the discrete solution.

C. Computational Complexity Analysis

Suppose we have a data matrix $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ and generate m anchors from $\boldsymbol{X}$. The steps of FSCAG and the corresponding computation complexity are summarized as follows.

1) We need $O(1)$ to obtain m anchors by random selection.
2) We need $O(ndm)$ to obtain the matrix $\boldsymbol{Z}$ to construct anchor graph.
3) We need $O(nmc+m^2c)$ to obtain the relaxed continuous solution of $\boldsymbol{F}$ by performming SVD on matrix $\boldsymbol{B}$.
4) We need $O(nmct)$ to perform k-means on the relaxed continuous solution for final clustering results, where t is the iterative number.

Considering that $m \ll n$, $c \ll d$, and t is usually small, the overall computational complexity of FSCAG is $O(ndm)$. We summarize the detail of our algorithm in Algorithm 1.

Algorithm 1 Algorithm of FSCAG

Input: HSI data matrix $X \in \mathbb{R}^{n\times d}$, the number of classes c, the number of anchors m, the tradeoff parameter α, the number of neighbors k.

1) Generate m anchors by random selction.
2) Obtain the matrix $\boldsymbol{Z}$ by (5).
3) Obtain the relaxed continuous solution of $\boldsymbol{F}$ by performing SVD on matrix $\boldsymbol{B}$, where $\boldsymbol{B} = \boldsymbol{Z}\boldsymbol{\Lambda}^{-\frac{1}{2}}$.
4) Perform k-means on the relaxed continuous solution of $\boldsymbol{F}$ for final clustering results.

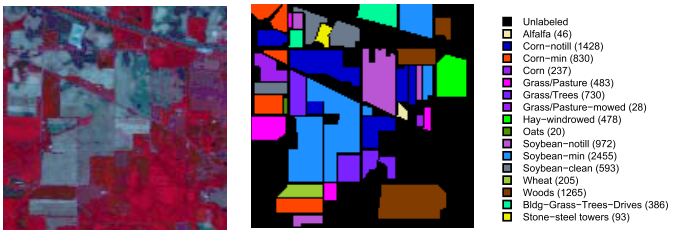
Output: c classes

III. Experimental Results

Here, we perform FSCAG on three widely used HSI data sets. In order to evaluate the performance of the FSCAG algorithm, k-means [4], FCM [5], FCM_S1 [6], and SC [11] were selected as benchmarks. The computational complexity of k-means, FCM, and FCM_S1 is $O(ndct)$, but that of SC is $O(n^2d)$. All our experiments are conducted on a Windows 10 computer with a 2.8 GHz Intel Xeon E3-1505M CPU and 32 GB RAM, MATLAB 2014a.

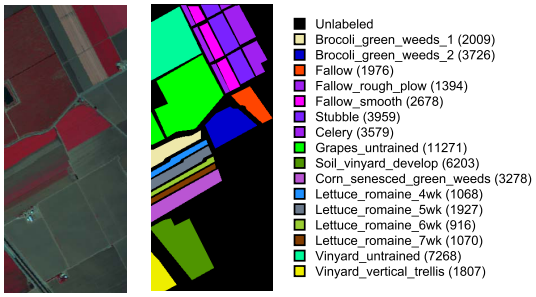
A. Hyperspectral Data Sets

1) *Indian Pines:* The data set was acquired by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor in 1992. The image scene contains 145 × 145 pixels and 220 spectral bands. A total of 20 water absorption and noisy bands (104–108, 150–163, and 220) were removed from the original 220 bands, leaving 200 spectral features for the experiment. The total number of samples is 21 025 and there are 16 classes in the data set. The false color image composition of bands 50, 27, and 17 and the ground truth map are provided in Fig. 1(a) and (b).



(a) (b)

Fig. 1. Indian Pines HI. (a) False color image (RGB 50, 27, 17). (b) Corresponding ground truth map. (Note that the number of each class is shown in brackets.)



(a) (b)

Fig. 2. Salinas HI. (a) False color image (RGB 70, 27, 17). (b) Corresponding ground truth map. (Note that the number of each class is shown in brackets.)

(a) (b)

Fig. 3. Pavia Center HI. (a) False color image (RGB 40, 30, 20). (b) Corresponding ground truth map. (Note that the number of each class is shown in brackets.)

2) *Salinas:* The data set was also acquired by the AVIRIS sensor. The image scene contains 512 × 217 pixels and 224 spectral bands. A total of 20 water absorption bands (108–112, 154–167, and 224) were removed from the original 224 bands, leaving 204 spectral features for the experiment. The total number of samples is 111 104 and there are 16 classes in the data set. The false color image composition of bands 70, 27, and 17 and the ground truth map are shown in Fig. 2(a) and (b).
3) *Pavia Center:* The data set was acquired by the Reflective Optics System Imaging Spectrometer sensor. The image scene contains 1096 × 1096 pixels and 102 spectral bands. Since some of the samples in the image contain no information and have to be discarded from the original image, leaving 1096 × 715 pixels for the experiment. The total number of samples is 783 640 and there are nine classes in the data set. The false color

TABLE I
QUANTITATIVE EVALUATIONS ON THREE HSI DATA SETS (OM ERROR)

Method \ Data set	Indian Pines			Salinas			Pavia Centre		
	AA	OA	Kappa	AA	OA	Kappa	AA	OA	Kappa
k-means	36.33	34.21	0.2689	63.63	66.97	0.6306	41.57	70.32	0.5917
FCM	37.40	36.35	0.2919	65.81	64.94	0.6132	44.71	71.77	0.6119
FCM_S1	34.56	37.20	0.3029	66.90	67.20	0.6333	47.00	73.24	0.6317
SC	30.66	33.45	0.2754	OM	OM	OM	OM	OM	OM
FSCAG	**49.53**	**41.21**	**0.3520**	**69.97**	**74.63**	**0.7173**	**64.94**	**81.55**	**0.7441**

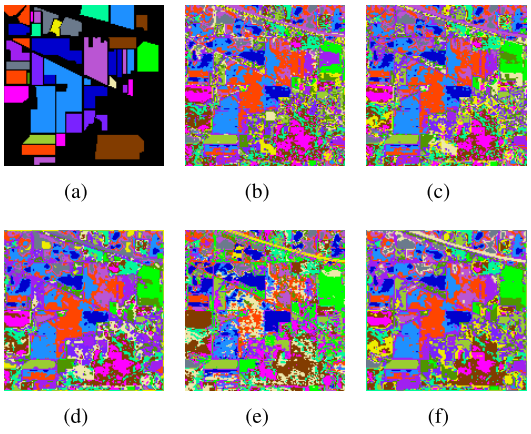


Fig. 4. Clustering maps on the Indian Pines data set. (a) Ground truth. (b) k-means. (c) FCM. (d) FCM_S1. (e) SC. (f) FSCAG.

image composition of bands 40, 30, and 20 and the ground truth map are shown in Fig. 3(a) and (b).

B. Clustering Results and Analysis

In our experiments, both quantitative evaluations (including average accuracy (AA), overall accuracy (OA), and kappa coefficient) and visual clustering maps are given to thoroughly evaluate the clustering performance of each clustering method. In addition, all the parameters of the compared clustering methods were manually tuned to the optimum. All the samples of the Indian Pines, Salinas, and Pavia Center data sets were selected as test data, including all classes and unlabeled pixels.

First, clustering experiments are conducted on the Indian Pines data set to evaluate the performance of each clustering method. The parameters used in FSCAG are set as $m = 500$, $k = 5$, and $\alpha = 0.6$, respectively. The quantitative evaluations and the clustering maps obtained with each clustering method are shown in Table I and Fig. 4, respectively. In Table I, the best results are shown in bold. From Table I and Fig. 4, we see that k-means, FCM, and SC obtain inferior clustering performances, but in contrast FCM_S1 and FSCAG improve the clustering performance to obtain much smoother clustering maps, which clearly reflects the importance of incorporating spatial information. Focusing on the running time of each clustering method, from Table II, we see that the running time of k-means, FCM, FCM_S1, and FSCAG is of the same order of magnitude. Especially for the two graph-based clustering methods, the FSCAG only needs 1.0 s, which is 54 times faster than the SC method. Note that SC can work on the Indian Pines data set that does not belong to large-scale HSI data set because the total number of samples is only 21 025.

TABLE II
RUNNING TIME ON THREE HSI DATA SETS (OM ERROR)

Dataset	k-means	FCM	FCM_S1	SC	FSC
Indian Pines	3.0s	4.5s	6.9s	54.3s	**1.0s**
Salinas	15.3s	24.0s	37.0s	OM	**11.9s**
Pavia Centre	142.2s	**90.9s**	138.0s	OM	134.1s

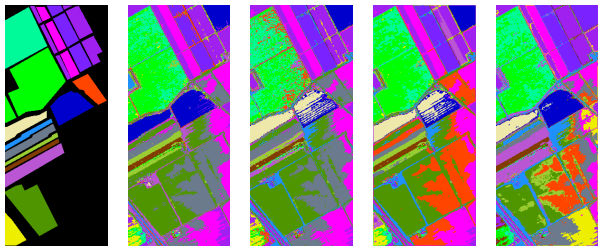
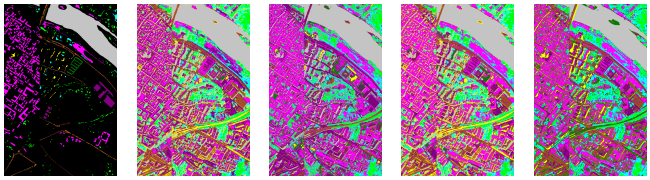


Fig. 5. Clustering maps on the Salinas data set. (a) Ground truth. (b) k-means. (c) FCM. (d) FCM_S1. (e) FSCAG.

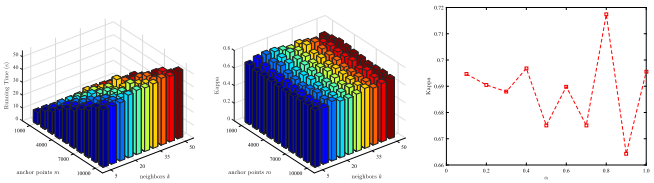
Then, clustering experiments are conducted on the Salinas data set. The parameters are set as $m = 1000$, $k = 5$, and $\alpha = 0.8$, respectively. The quantitative evaluations and the clustering maps obtained with each clustering method are shown in Table I and Fig. 5, respectively. By referring to Table I, FSCAG obtains the highest precision with the best OA of 74.63% and a kappa coefficient of 0.7173. From Fig. 5, we see that FSCAG produces more homogenous areas and better clustering maps than other algorithms. Table II shows that the running time of k-means, FCM, FCM_S1, and FSCAG is of the same order of magnitude. Note that the Salinas data set belongs to large-scale HSI data set and the total number of samples is 111 104. Focus on the two graph-based clustering methods, the FSCAG only needs 11.9 s, but SC cannot work on the Salinas data set due to "out-of-memory (OM) error."

Finally, clustering experiments are conducted on the Pavia Center data set. The parameters are set as $m = 1000$, $k = 5$, and $\alpha = 0.3$, respectively. The quantitative evaluations obtained with each clustering method are given in Table I, and the corresponding clustering maps are shown in Fig. 6. By referring to Table I, FSCAG obtains the highest precision with the best OA of 81.55% and a kappa coefficient of 0.7441. From Fig. 6, we see that FSCAG produces more homogenous areas and better clustering maps than other algorithms. Moreover, FCM_S1 and FSCAG obtain better clustering results and smoother clustering maps by considering the spatial information. From Table II, we see that the running time of k-means, FCM, FCM_S1, and FSCAG is 142.2, 90.9, 138.0, and 134.1, respectively. As the growth of the scale of Pavia



(a) (b) (c) (d) (e)

Fig. 6. Clustering maps on the Pavia Center data set. (a) Ground truth. (b) k-means. (c) FCM. (d) FCM_S1. (e) FSCAG.



(a) (b) (c)

Fig. 7. Parameter sensitivity experiments on the Salinas data set. (a) Running time with different m and k values. (b) Kappa with different m and k values. (c) Kappa with different α values.

Center data set, the total number of samples increases to 783 640. Focus on the two graph-based clustering methods, the FSCAG needs 134.1 s, but SC cannot work on the Pavia Center data set due to "OM error."

C. Parameter Sensitivity

In this section, we evaluate the parameter sensitivity for our proposed FSCAG algorithm. There are three parameters that should be tuned: m, k, and α. From Section II-C, we see that the computational complexity is mainly determined by the parameter m. The parameter k is related to the sparsity of the matrix $\boldsymbol{Z}$ and has little effect on the computational complexity. In addition, the parameter α has no relevance to the computational complexity.

We conduct parameter sensitivity experiments on the Salinas data set. By referring to Fig. 7(a), we can see that the running time depends mainly on the parameter m, and the parameter k has little effect on the running time. Fig. 7(b) shows that the clustering performance of our FSCAG algorithm is relatively robust to the the parameters m and k. Noting that the motivation of our FSCAG algorithm is fast, it is reasonable to select small $m = 1000$ and $k = 5$ for the Salinas data set.

Based on this, we only need to tune the parameter α. Fig. 7(c) shows that our FSCAG algorithm is sensitive to the parameter α. It is still an open problem to decide the value of the parameter α for different data sets.

IV. Conclusion

In this letter, we have presented a novel FSCAG algorithm to efficiently deal with the large-scale HSI clustering problem. By incorporating the spatial information of HSI, the anchor graph is first constructed. Then, spectral analysis is performed on the anchor graph. The overall computational complexity of our FSCAG algorithm reduces to $O(ndm)$, which is a significant improvement compared to conventional graph-based clustering that needs $O(n^2d)$. Experimental results clearly show that the FSCAG algorithm achieves good clustering performance for large-scale HSI.

Acknowledgment

The authors would like to thank Prof. D. Landgrebe at Purdue University for providing the free downloads of the Indian Pines and Salinas AVIRIS images and Prof. Gamba from the University of Pavia for providing the Pavia Center data set.

References

[1] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, Jun. 2014.

[2] B. Sun, X. Kang, S. Li, and J. A. Benediktsson, "Random-walker-based collaborative learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 212–222, Jan. 2017.

[3] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016.

[4] J. A. Hartigan and M. A. Wong, "A k-means clustering algorithm," *Appl. Stat.*, vol. 28, no. 1, pp. 100–108, 1979.

[5] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum, 1981.

[6] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.

[7] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.

[8] S. Vijendra, "Efficient clustering for high dimensional data: Subspace based clustering and density based clustering," *Inf. Technol. J.*, vol. 10, no. 6, pp. 1092–1105, May 2011.

[9] Y. Zhong, L. Zhang, and W. Gong, "Unsupervised remote sensing image classification using an artificial immune network," *Int. J. Remote Sens.*, vol. 32, no. 19, pp. 5461–5483, Oct. 2011.

[10] Y. Zhong, S. Zhang, and L. Zhang, "Automatic fuzzy clustering based on adaptive multi-objective differential evolution for remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2290–2301, Oct. 2013.

[11] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. NIPS*, 2002, pp. 849–856.

[12] G. Chen and G. Lerman, "Spectral curvature clustering (SCC)," *Int. J. Comput. Vis.*, vol. 81, no. 3, pp. 317–330, Mar. 2009.

[13] H. Zhai, H. Zhang, L. Zhang, P. Li, and A. Plaza, "A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 43–47, Jan. 2017.

[14] W. Liu, J. He, and S. F. Chang, "Large graph construction for scalable semi-supervised learning," in *Proc. ICML*, 2010, pp. 679–686.

[15] C. Deng, R. Ji, W. Liu, D. Tao, and X. Gao, "Visual reranking through weakly supervised multi-graph learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2600–2607.

[16] C. Deng, R. Ji, D. Tao, X. Gao, and X. Li, "Weakly supervised multi-graph learning for robust image reranking," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 785–795, Apr. 2014.

[17] D. Cai and X. Chen, "Large scale spectral clustering via landmark-based sparse representation," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1669–1680, Aug. 2015.

[18] F. Nie, W. Zhu, and X. Li, "Unsupervised large graph embedding," in *Proc. AAAI*, 2017, pp. 2422–2428.

[19] F. Nie, X. Wang, M. Jordan, and H. Huang, "The constrained Laplacian rank algorithm for graph-based clustering," in *Proc. AAAI*, 2016, pp. 1969–1976.

[20] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. Farag, and T. Moriarty, "A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, no. 3, pp. 193–199, Mar. 2002.

[21] R. Zass and A. Shashua, "Doubly stochastic normalization for spectral clustering," in *Proc. NIPS*, 2007, pp. 1569–1576.