

# LANDMARK-BASED LARGE-SCALE SPARSE SUBSPACE CLUSTERING METHOD FOR HYPERSPECTRAL IMAGES

*Shaoguang Huang<sup>1</sup>, Hongyan Zhang<sup>2</sup> and Aleksandra Pižurica<sup>1</sup>*

<sup>1</sup>Department of Telecommunications and Information Processing, TELIN-IPI-imec,  
Ghent University, Belgium

<sup>2</sup>The State Key Lab. of Inform. Engineering in Surveying, Mapping, and Remote Sensing,  
Wuhan University, China

## ABSTRACT

Sparse subspace clustering (SSC) has achieved the state-of-the-art performance in the clustering of hyperspectral images (HSIs). However, the high computational complexity and sensitivity to noise limit its clustering performance. In this paper, we propose a scalable SSC method for the large-scale HSIs, which significantly accelerates the clustering speed of SSC without sacrificing clustering accuracy. A small landmark dictionary is first generated by applying k-means to the original data, which results in the significant reduction of the number of optimization variables in terms of sparse matrix. In addition, we incorporate spatial regularization based on total variation (TV) and improve this way strongly robustness to noise. A landmark-based spectral clustering method is applied to the obtained sparse matrix, which further improves the clustering speed. Experimental results on two real HSIs demonstrate the effectiveness of the proposed method and the superior performance compared to both traditional SSC-based methods and the related large-scale clustering methods.

**Index Terms**— Sparse subspace clustering, landmark, hyperspectral image, large-scale data

## 1. INTRODUCTION

Due to their rich spectral information, hyperspectral images (HSIs) find numerous applications including remote sensing for defence and security, agricultural monitoring and geosciences. Clustering of HSIs, as a fundamental task, is to discriminate the pixels to different groups corresponding to different materials or objects without any prior label information.

In recent years, sparse subspace clustering (SSC) [1] method has achieved the state-of-the-art performance in HSI clustering [2–7]. SSC is based on a self-representation model

which employs the input data as a dictionary. With a sparsity constraint, SSC claims a subspace-preserving property that for a given sample from a union of subspaces there exist a sparse representation vector whose non-zero entries correspond to the samples in the same subspaces. Therefore, SSC can automatically choose the nearby data points for each input data point. Then similarity matrix is constructed by using the sparse coefficient matrix directly. In the end, the clustering results can be obtained by applying similarity matrix into the standard spectral clustering framework. However, two important problems limit severely the applicability of SSC in real applications. The first limitation is huge sensitivity to noise. This is because SSC treats each pixel in HSIs independently in the sparse representation model. The second limitation is the enormous computational complexity when dealing with large images, which are common in HSIs. The time complexity of SSC reaches more than  $\mathcal{O}(m^3)$  in one iteration where  $m$  is the number of pixels. Hence, for typical HSIs, SSC model is often infeasible despite its superior theoretical performance.

To reduce the noise sensitivity, some extensions of SSC have been proposed by introducing different spatial constraints [2–7]. A smoothing strategy was introduced in [2] by minimizing the coefficient difference between the central pixel and the mean of pixels in a local square window. A  $\ell_2$  norm based spatial regularization was employed in [4] to smooth the coefficients in the small local region. In [5, 7], a joint sparsity constraint with  $\ell_{1,2}$  norm was applied to the coefficient of each super-pixel. The work in [6] exploited the collaborative representation with a  $\ell_2$  norm and proposed a total variation (TV) regularized model with a locally adaptive dictionary. However, those traditional SSC-based methods are not scalable, and thereby cannot be applied on the large-scale data sets.

Some general large-scale clustering methods based on SSC have been proposed in the area of computer vision to cope with the computational complexity problem [8, 9]. The work in [8] applies SSC only to a small number of pre-selected data points and using the clustering result on this

---

This work was supported in part by the Fonds voor Wetenschappelijk Onderzoek (FWO) project: G.OA26.17N and in part by the National Natural Science Foundation of China under grants 61871298 and 41711530709.

small set, constructs a dictionary for the sparse representation classification (SRC), which is applied to the rest of the data. In [9], a sketched SSC model was proposed by using a random projection technique, which compressed the large dictionary to a much smaller one. Those methods are feasible for the large-scale HSIs, however, they usually obtain lower clustering accuracy when noise is present in the data.

In order to address the problems mentioned above, we propose a landmark-based SSC model with TV regularization (LSSC-TV) for the large-scale HSIs. Instead of the self-representation dictionary, we select some representative pixels as the landmarks to construct the dictionary which has a much smaller size, resulting in the significant reduction of the number of optimization variables. In addition, a spatial regularization based on the TV-norm is incorporated in the proposed model, which improves greatly the robustness to noise. Moreover, a non-negative constraint and a sum-to-one constraint are added to interpret the value of coefficients as the probabilities for a pixel to choose the landmarks. After obtaining the sparse matrix, we calculate the final clustering results by applying the landmark-based spectral clustering (LSC) method [10].

The rest of the paper is organized as follows, Section 2 briefly introduces the clustering of HSIs with SSC model. Section 3 describes the proposed LSSC-TV model. Section 4 presents the experimental results on real data and Section 5 concludes the paper.

## 2. THE SSC MODEL OF HSI CLUSTERING

We denote by  $\mathbf{Y} \in \mathbb{R}^{B \times MN}$  the flattened 2-D matrix from the original 3-D HSI data cube with a size of  $M \times N \times B$ , where  $M$  and  $N$  represent the height and the width of the HSI, respectively, and  $B$  denotes the number of bands. Let  $c$  be the number of classes. With the self-representation dictionary  $\mathbf{Y}$ , the SSC model [1] with respect to the sparse matrix  $\mathbf{C} \in \mathbb{R}^{MN \times MN}$  is derived by:

$$\begin{aligned} \arg \min_{\mathbf{C}} \|\mathbf{C}\|_1 + \frac{\lambda}{2} \|\mathbf{Y} - \mathbf{Y}\mathbf{C}\|_F^2 \\ \text{s.t. } \text{diag}(\mathbf{C}) = \mathbf{0}, \mathbf{1}^T \mathbf{C} = \mathbf{1}^T, \end{aligned} \quad (1)$$

where  $\|\mathbf{C}\|_1 = \sum_i \sum_j |C_{ij}|$ ;  $\mathbf{1}$  is an all-one vector;  $\text{diag}(\mathbf{C})$  extracts the diagonal elements of  $\mathbf{C}$ ;  $\lambda$  is a parameter, which controls the balance between the data fidelity and the sparsity of the coefficient matrix. The first constraint is introduced to avoid the trivial solution of representing a sample by itself and the second constraint indicates the case of affine subspace.

The model in (1) can be solved by ADMM [11]. Due to the subspace preserving property of SSC, the sparse matrix  $\mathbf{C}$  can be used to construct the similarity matrix  $\mathbf{W} \in \mathbb{R}^{MN \times MN}$  by  $\mathbf{W} = |\mathbf{C}| + |\mathbf{C}|^T$ . Then the clustering results can be obtained by applying  $\mathbf{W}$  in the spectral clustering [12]. Specifically, the  $c$  eigenvectors of Laplacian matrix  $\mathbf{L} = \text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}$  corresponding to the  $c$  smallest

eigenvalues of  $\mathbf{L}$  are first calculated by singular-value decomposition (SVD), and then applied to the k-means clustering method.

## 3. PROPOSED LSSC-TV MODEL

SSC and the traditional SSC-based methods [2–7] suffer from the high computational complexity problem, resulting from the iterative optimization. A key obstacle is that the inverse of the entire large matrix  $(\mathbf{Y}^T \mathbf{Y} + \mu \mathbf{I}) \in \mathbb{R}^{MN \times MN}$  has to be calculated and saved in memory, which results in time complexity of  $\mathcal{O}((MN)^3)$ . As the value of  $MN$  in HSIs is usually very large, such methods impose huge requirements in terms of computational resources and may require tremendous running time, which means they are not scalable. Another limitation of SSC is its sensitivity to noise, which is a consequence of the independent sparse coding, which ignores the spatial continuity of pixel values in a local region. The inferred similarity matrix is thus prone to noise and large spectral variations within-class, which deteriorates thereby the spectral clustering performance.

In order to reduce the computation complexity, we replace the self-representation dictionary  $\mathbf{Y}$  with a small landmark-based dictionary  $\mathbf{D} \in \mathbb{R}^{B \times n}$  and represent the input data as a linear combination of those landmarks which are generated by k-means in the original data  $\mathbf{Y}$ . This way we significantly reduce the number of optimization variables from  $(MN)^2$  to  $MNn$ , where  $n \ll MN$ .

Next, we reduce the noise sensitivity by introducing an appropriate regularization. Observe that pixels in HSIs are spatially correlated and neighbouring pixels belong to the same class with high probability. Therefore, the coefficients of neighbouring pixels should be similar. To account for these spatial properties, we impose TV-norm regularization on the coefficients matrix. The coefficients in the proposed method are interpreted as the probability to select landmarks, thus non-negative and sum-to-one constraints are added for this physical meaning. The proposed LSSC-TV model with respect to sparse matrix  $\mathbf{A} \in \mathbb{R}^{n \times MN}$  is derived by:

$$\begin{aligned} \arg \min_{\mathbf{A}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_1 + \lambda_{tv} \|\mathbf{A}\|_{TV} \\ \text{s.t. } \mathbf{A} \geq 0, \mathbf{1}^T \mathbf{A} = \mathbf{1}^T \end{aligned} \quad (2)$$

where  $\lambda$  and  $\lambda_{tv}$  are the penalty parameters for the sparsity level and spatial smoothness, respectively, and the TV norm is defined by:

$$\|\mathbf{A}\|_{TV} = \|\mathbf{H}_x \mathbf{A}^T\|_1 + \|\mathbf{H}_y \mathbf{A}^T\|_1, \quad (3)$$

where  $\mathbf{H}_x$  and  $\mathbf{H}_y$  are the forward finite-difference operators in the horizontal and vertical directions, respectively, with periodic boundary conditions.

The proposed model can be efficiently solved by the ADMM algorithm [11]. After obtaining the sparse matrix,

the next step is to build similarity matrix. As sparse matrix  $\mathbf{A}$  here is also a probability matrix indicating how likely a data point selects the landmarks  $\mathbf{d}_i (i = 1, 2, \dots, n)$ , based on the theory of AnchorGraph in [13], we can construct the similarity matrix directly by:

$$\mathbf{W} = \mathbf{A}^T \mathbf{\Lambda}^{-1} \mathbf{A} = \hat{\mathbf{A}}^T \hat{\mathbf{A}}, \quad (4)$$

where  $\mathbf{\Lambda} \in \mathbb{R}^{n \times n}$  is a diagonal matrix with  $\Lambda_{i,i} = \sum_{j=1}^{MN} A_{i,j}$  and  $\hat{\mathbf{A}} = \mathbf{\Lambda}^{-1/2} \mathbf{A}$ .

In the case that  $\mathbf{W}$  is sparse, less memory will be needed to save the matrix and we can directly apply  $\mathbf{W}$  within the spectral clustering method. However, the sparseness of  $\mathbf{W}$  is not guaranteed, which may suffer from the high requirements in terms of memory needed to save the large dense matrix  $\mathbf{W}$  for the large-scale data. Inspired by the landmark-based spectral clustering method [10], we can calculate the first  $c$  eigenvectors of  $\hat{\mathbf{A}} \hat{\mathbf{A}}^T \in \mathbb{R}^{n \times n}$ , denoted by  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c]$ , corresponding to eigenvalues  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_c$ , and apply the matrix  $\mathbf{\Sigma}^{-1} \mathbf{V}^T \hat{\mathbf{A}}$  to the k-means to obtain the clustering results, where  $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_c) \in \mathbb{R}^{c \times c}$ . Note that the value of  $n$  is far smaller than  $MN$ , therefore, the eigenvectors of  $\hat{\mathbf{A}} \hat{\mathbf{A}}^T$  can be calculated very fast. We refer to the paper [10] for details on the clustering method.

## 4. EXPERIMENTAL RESULTS

We conduct experiments on two widely used benchmark data sets: *Indian Pines* and *Salinas*. The results of the original SSC [1], spatial-regularized L2-SSC [4] and the large-scale clustering methods SSSC [8], LSC [10] and Sketch-SSC [9] are reported for comparison. Three performance measures: overall accuracy (OA), Kappa coefficient ( $\kappa$ ) and running time ( $t$ ) are used for quantitative assessment. All the methods are implemented in MATLAB on a computer with an Intel<sup>®</sup> core-i7 3930K CPU with 64 GB of RAM. The results of SSSC, LSC, Sketch-SSC and LSSC-TV are reported in average of 5 runs.

### 4.1. Results with small data set

As the traditional SSC-based methods [1, 4] pose excessive memory requirements in case of large-scale data sets, we first test all the methods on the cropped *Indian Pines*, which is commonly used in the literature [2, 5, 7]. The cropped data set has 4 classes and the image size is  $85 \times 70 \times 200$ . The parameters of the proposed method LSSC-TV are set as  $\lambda = 5 \times 10^{-3}$ ,  $\lambda_{tv} = 1 \times 10^{-2}$  and  $n = 500$ , empirically. The clustering results reported in Table 1 demonstrate that the proposed LSSC-TV method achieves the highest clustering accuracy of 87.36%. Compared with the traditional SSC-based methods SSC and L2-SSC, LSSC-TV obtains nearly 20% accuracy improvement and 14 times speed up, which mainly benefits from the employed spatial regularization and the compact dictionary constructed by landmarks. Compared with the large-scale clustering methods SSSC, LSC and Sketch-SSC, the

**Table 1.** Clustering results for the cropped *Indian Pines*.

Class name	SSC	L2-SSC	SSSC	LSC	Sketch-SSC	LSSC-TV
Corn-notill	60.00	61.09	56.98	57.01	<b>62.19</b>	49.40
Grass-trees	98.36	99.32	96.66	96.49	<b>100</b>	99.95
Soybean-notill	76.91	79.37	70.30	55.66	68.80	<b>99.27</b>
Soybean-mintill	50.68	54.89	56.65	61.07	58.87	<b>97.88</b>
OA(%)	65.11	67.78	65.65	65.13	68.12	<b>87.36</b>
$\kappa$	0.5296	0.5629	0.5296	0.5119	0.5628	<b>0.8169</b>
$t(s)$	543	624	5	<b>0.4</b>	3	38

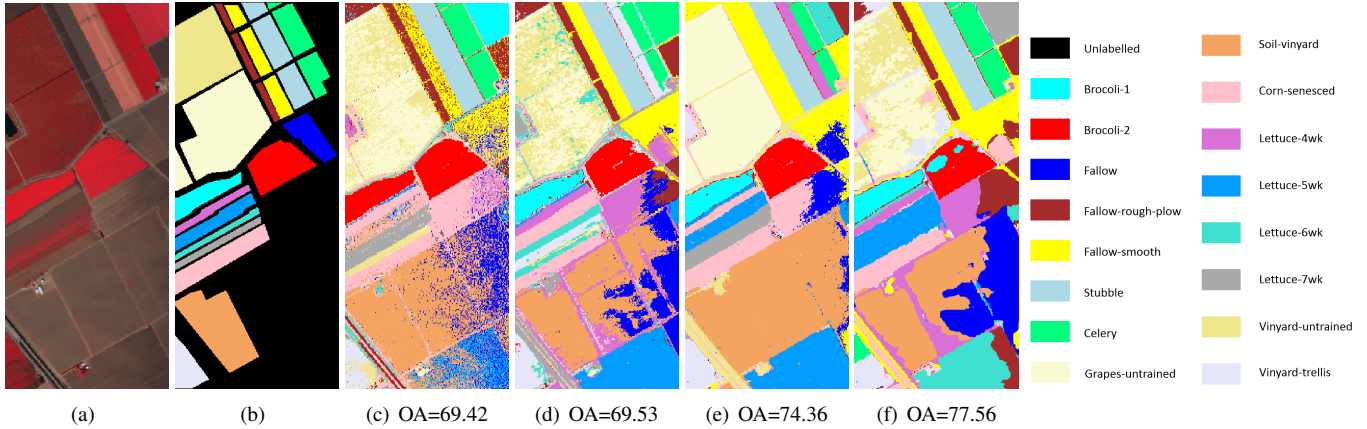
proposed method also yields significant accuracy improvement with comparable running time.

### 4.2. Results with large-scale data set

We also test the performance on a larger data set: *Salinas*, with size  $512 \times 217 \times 204$ , including 111,104 pixels. There are 16 classes in total. The false color and ground truth are shown in Fig. 1 (a) and (b). The parameters of our method are empirically set as  $\lambda = 1 \times 10^{-3}$ ,  $\lambda_{tv} = 5 \times 10^{-4}$  and  $n = 1000$ . Due to the excessive memory requirements of SSC and L2-SSC, they cannot be run on our computer for this data. The results of large-scale methods reported in Table 2 indicate that LSSC-TV consistently obtains the highest accuracy. Note that LSSC-TV and LSC use the same spectral clustering method and the difference is in the way to obtain  $\mathbf{A}$ . LSC employs  $k$  nearest neighbours (knn) method to find the closer landmarks, while LSSC-TV automatically selects the nearby landmarks based on sparse representation. In contrast with knn, the sparsity-based model is more flexible in the embedding of various spatial regularizations. In terms of running time, LSC is the fastest method, and our method obtains comparable running time with SSSC and Sketch-SSC. The clustering maps in Fig. 1 show that the proposed method and Sketch-SSC suffer from less impulse noise than SSSC and LSC in this data.

## 5. CONCLUSION

In this paper, a scalable SSC model was proposed by incorporating the landmark-based sparse representation, which is feasible for both small and large-scale HSIs without sacrificing clustering accuracy. The small landmark dictionary reduces the amount of optimization variables, resulting in a scalable clustering model. The exploitation of spatial regularization based on TV norm strengthens the dependency between neighbouring pixels and guarantees thereby the robustness to noise and spectral diversity within-class. Moreover, landmark-based spectral clustering method is introduced to



**Fig. 1.** *Salinas*. (a) False color image, (b) Labels, and clustering maps of (c) SSSC, (d) LSC, (e) Sketch-SSC and (f) LSSC-TV.

**Table 2.** Clustering results for *Salinas*.

Class name	SSSC	LSC	Sketch-SSC	LSSC-TV
Brocoli-1	59.06	78.52	<b>99.43</b>	40.00
Brocoli-2	98.34	97.20	<b>98.91</b>	92.56
Fallow	5.75	2.58	11.92	<b>12.42</b>
Fallow-rough-plow	77.45	99.24	19.90	<b>99.33</b>
Fallow-smooth	89.80	99.10	<b>99.45</b>	98.96
Stubble	<b>99.61</b>	98.96	99.54	98.84
Celery	90.82	69.72	55.25	<b>99.61</b>
Grapes-untrained	69.08	47.90	<b>98.67</b>	66.58
Soil-vinyard	97.04	92.09	<b>99.72</b>	91.46
Corn-senesced	43.00	64.21	88.04	<b>90.17</b>
Lettuce-4wk	73.01	<b>97.62</b>	54.21	92.53
Lettuce-5wk	14.98	0	70.97	<b>88.57</b>
Lettuce-6wk	0	<b>97.84</b>	0	0
Lettuce-7wk	90.02	0.04	<b>97.78</b>	0
Vinyard-untrained	48.30	59.91	0.28	<b>63.71</b>
Vinyard-trellis	46.69	61.42	<b>97.61</b>	74.47
OA	68.81	67.10	73.43	<b>75.40</b>
$\kappa$	0.6515	0.6387	0.7007	<b>0.7273</b>
$t(s)$	116	<b>21</b>	269	353

solve the dense similarity matrix problem for the large-scale HSIs.

## 6. REFERENCES

- [1] E. Elhamifar and R. Vidal, “Sparse subspace clustering: Algorithm, theory, and applications,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, 2013.
- [2] 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, 2016.
- [3] H. Zhai, H. Zhang, X. Xu, L. Zhang, and P. Li, “Kernel sparse subspace clustering with a spatial max pooling operation for hyperspectral remote sensing data interpretation,” *Remote Sensing*, vol. 9, no. 4, pp. 335, 2017.
- [4] 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, 2017.
- [5] S. Huang, H. Zhang, and A. Pižurica, “Joint sparsity based sparse subspace clustering for hyperspectral images,” in *Proc. IEEE ICIP*, 2018, pp. 3878–3882.
- [6] H. Zhai, H. Zhang, L. Zhang, and P. Li, “Total variation regularized collaborative representation clustering with a locally adaptive dictionary for hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 166–180, 2018.
- [7] S. Huang, H. Zhang, and A. Pižurica, “Semi-supervised sparse subspace clustering method with a joint sparsity constraint for hyperspectral remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, under review, 2018.
- [8] X. Peng, L. Zhang, and Z. Yi, “Scalable sparse subspace clustering,” in *Proc. IEEE CVPR*, 2013, pp. 430–437.
- [9] P. A. Traganitis and G. B. Giannakis, “Sketched subspace clustering,” *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1663–1675, 2018.
- [10] D. Cai and X. Chen, “Large scale spectral clustering via landmark-based sparse representation,” *IEEE Transactions on Cybernetics*, vol. 45, no. 8, pp. 1669–1680, 2015.
- [11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [12] U. Von Luxburg, “A tutorial on spectral clustering,” *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [13] W. Liu, J. He, and S. Chang, “Large graph construction for scalable semi-supervised learning,” in *Proc. ICML*, 2010, pp. 679–686.