Spatial-spectral Based Multi-view Low-rank Sparse Sbuspace Clustering for Hyperspectral Imagery

Long Tian [1], *Qian Du* [1], *Ivica Kopriva* [2], *Nicolas Younan* [1]

[1] Department of Electrical and Computer Engineering, Mississippi State University, USA
[2] Division of Electronics, Ruđer Bošković Institute, Croatia

***Abstract*—Hyperspectral image (HSI) Clustering is an unsupervised task, which segments pixels into different groups without using labeled samples. In this paper, spatial-spectral based multi-view low-rank sparse subspace clustering (SSMLC) algorithm is proposed. Due to significant number of spectra bands HSI contains much more information than a regular image. These spectral information can be considered as multi-view. In this paper, the spectral partitioning is applied to generate spectral views which contain correlated bands. Morphological features of the original HSI are taken as another view which contains spatial features. Principal components construct another view, which eliminates the noise in the original dataset. After the multi-view dataset is formed, multi-view low-rank sparse subspace clustering is applied to segment HSI. Our experiments show that the performance of the proposed SSMLC is better than other that of clustering algorithms such as sparse subspace clustering and low-rank sparse subspace clustering.**

***Keywords* — Hyperspectral image, clustering, low-rank sparse subspace clustering, multi-view learning, spatial-spectral feature**

I. INTRODUCTION

By combining spatial and spectral information, a hyperspectral image (HSI) provides much more information than a regular image. Because of this particular advantage, HSI has been widely applied in remote sensing monitoring [1-5], where its high spectral resolution can help distinguish materials with subtle spectral discrepancy. Clustering is one of popular techniques in image processing. For an HSI, it separates pixels into corresponding groups by considering both spectral and spatial information. As an unsupervised technique, clustering is more challenging than supervised ones using labeled samples.

The *k*-means clustering is a classical method, which heavily relies on initial conditions and easily gets stuck in a local optimum. Furthermore, the clustering results are centroid-based, but the HSI do not have this nature. In this case, the subspace clustering algorithm is proposed for high-dimensional dataset, where the data are clustered into multiple subspaces. Afterwarrds, low-dimensional subspaces are fitted to each group of pixels. Recently, sparse subspace clustering (SSC) and low-rank subspace clustering (LSC) [7,8] are proposed to find affinity matrices for clustering effectively, where an affinity matrix defines the similarity between pixels. The SSC algorithm uses the sparsest representation for each pixel with pixels in its group, and the local structure of data can be maintained. The LRR algorithm introduces low-rank constraint into self representation matrix, and the global structure of data is preserved. In order to contain both local and global information in dataset, the low-rank sparse subspace clustering (LRSSC) algorithm is proposed which combines the low-rank and sparsity constraints [9].

Because of varied sources or features of dataset, mult-view learning has been widely applied in machine learning area, where the single-view learning could not represent all the sources or features in dataset properly. HSI, which contains hundreds of spectral bands, could be represented with many views in each band. In this case, HSI is a perfect dataset for multi-view learning. Li *et al*. also utilized multiple morphological features for HSI classification [10], which integrates linear and nonlinear features.

Multi-view learning and LRSSC are incorporated in Ref. [11] as multi-view low-rank sparse subspace clustering (MLRSSC) to deal with multi-source or multi-features in dataset. In this research, we propose the use of MLRSSC for HSI clustering. In order to construct a multi-view data for such a task, spectral partitioning is applied to generate multi-views for spectral information. Afterwards, morphological component analysis is applied to produce another view of HSI to represent spatial information, while PCA is applied to yield an additional view of the original dataset. Finally, those views are treated as multi-source for MLRSSC clustering, which is named as spatial-spectral based multi-view low-rank sparse subspace clustering (SSMLC).

II. PROPOSED METHOD

A. Low-rank representation and sparse subspace clustering

Given a set of N data points as $X=[x_1,x_2\cdots x_N]$ in R^D, the low-rank representation matrix $Z\in R^{N\times N}$ could be recovered by the following minimization problem:

$$\min_{Z,E} rank(Z)+\gamma\|E\|_\ell,\ s.t.\ X=AZ+E \tag{1}$$

where A is a "dictionary" that linearly spans the whole data space, E is the error. In the LRSSC problem, the self representation is used, where $A=X$. In this case, equation (1) is simplified as:

$$\min_{Z}\|Z\|_*,\ \ s.t.\ X=XZ \tag{2}$$

where the rank of Z is approximated by nuclear norm of Z. When there is no noise in X or noise can be ignored, Z could be achieved by:

$$Z = VV^T \tag{3}$$

where V is the SVD components of X as $U\Sigma V^T$ [9].

Instead of low-rank representation where the whole data space is applied as representation space, SSC uses a small number of data points from its subspace for representation, which contains more local information in dataset. The SSC minimization problem is:

$$\min_Z \|Z\|_1, \ s.t. \ \text{X=XZ}, \ diag(Z) = 0. \tag{4}$$

where ℓ_1 norm regularization is used as the faithful representation from its own subspace [12-14], and $diag(Z) = 0$ constraint is applied to exclude a trivial solution where the data pixels are a linear combination of themselves. To solve minimization problem in (4), the alternating direction method of multipliers (ADMM) optimization algorithm is an efficient candidate.

B. Low-rank sparse subspace clustering

According to [15], sparse representation contains major local structure information of dataset, where each pixel has the sparsest representation. On the other hand, LRR focuses on the global structure information of the dataset. In this case, the combination of SSC and LRR algorithm, which is low-rank sparse subspace clustering (LRSSC) proposed by [9], could extract more information from the original dataset, and handle both global and local structure information. The minimization problem for the LRSSC can be expressed as

$$\min_Z \alpha_1 \|Z\|_* + \alpha_2 \|Z\|_1, \ s.t. \ \text{X=XZ}, \ \text{diag}(Z) = 0 \tag{5}$$

The affinity matrix W iscalculated as:

$$W = |Z| + |Z|^T \tag{6}$$

Then spectral clustering [17] is then applied to achieve clustering.

C. Spetral-spatial based multi-view low-rank sparse subspace clustering (SSMLC)

With the development of data acquisition technology, there are plenty of data sources or data features for a single object or event, and an individual view (source or feature) is not comprehensive for data description. In this case, multi-view learning algorithms, which integrate multiple sources or features, is a popular and successfully method in computer vision and intelligent system area [16]. In hyperspectral imaging, multi-view learning has been applied in classification as multiple features of HSI are combined together [10]. They showed multi-view learning could provide more information than single-view learning, yielding significant classification improvement.

Intuitively, multi-view learning could improve the performance of unsupervised clustering. For the MLRSSC, let *t* views construct a dataset $X = [X^1, X^2 \cdots X^t]$, where the *i*-th view $X^i = \{x_j^i\}_{j=1}^N \in R^{Di}$ contains D^i dimension features. The joint optimization problem with t views is:

$$\min_{Z^1, Z^2, \cdots Z^t} \sum_{i=1}^{t} \left(\alpha_1 \|Z^i\|_* + \alpha_2 \|Z^i\|_1 \right) + \lambda \sum_{1 \le i,j \le t, t \ne j} \|Z^i - Z^j\|_F^2,$$

$$s.t. \ X^i = X^i Z^i, diag(Z^i) = 0 \tag{7}$$

where the weight of each view λ could be simply assumed identical. In addition to the low-rank and sparsity constraints, the third term in equation (7) encourages the representations from different views to be consistent if possible.

By fixing all but one Z^t, equation (7) can be minimized for each Z^t independently, which can be reformulated as:

$$\min_{Z^t} \alpha_1 \|Z^t\|_* + \alpha_2 \|Z^t\|_1 + \lambda \sum_{j \ne t} \|Z^t - Z^j\|_F^2,$$

$$s.t. \ X^t = X^t Z^t, diag(Z^t) = 0 \tag{8}$$

By introducing auxiliary variables Z_1^t, Z_2^t, Z_3^t, A^t, equation (8) becomes

$$\min_{Z_1^t, Z_2^t, Z_3^t, A^t} \alpha_1 \|Z_1^t\|_* + \alpha_2 \|Z_2^t\|_1 + \lambda \sum_{j \ne t} \|Z_3^t - Z^j\|_F^2,$$

$$s.t. \ X^t = X^t A^t, A^t = Z_2^t - diag(Z_2^t), A^t = Z_1^t, A^t = Z_3^t \tag{9}$$

The augmented Lagrangian can be formed as:

$$L\left(\{Z_i^t\}_{i=1}^3, A^t, \{\Lambda_i^t\}_{i=1}^4\right) = \alpha_1 \|Z_1^t\|_* + \alpha_2 \|Z_2^t\|_1 + \lambda \sum_{j \ne t} \|Z_3^t - Z^j\|_F^2$$
$$+ \frac{\mu_1}{2} \|X^t - X^t A^t\|_F^2 + \frac{\mu_2}{2} \|A^t - Z_2^t + diag(Z_2^t)\|_F^2 + \frac{\mu_3}{2} \|A^t - Z_1^t\|_F^2$$
$$+ \frac{\mu_4}{2} \|A^t - Z_3^t\|_F^2 + tr[\Lambda_1^{t\,T}(X^t - X^t A^t)] + tr[\Lambda_3^{t\,T}(A^t - Z_1^t)]$$
$$+ tr[\Lambda_4^{t\,T}(A^t - Z_3^t)] + tr[\Lambda_2^{t\,T}(A^t - Z_2^t + diag(Z_2^t))] \tag{10}$$

where $\{\Lambda_i^t\}_{i=1}^4$ are Lagrange dual parameters and μ's are penalty variables. Then the ADMM can be applied to solve this convex optimization problem.

Finally, spectral clustering [17] is applied to the affinity matrix *W* which is generated from *Z* using equation (6).

In this research, spectral partitioning based on correlation coefficients is applied to generate multiple spectral views, and a highly correlated spectral group is considered as a view. Principal component analysis (PCA) is also applied to decorrelate the data and remove noise, and the principal components are considered as a view. To generate spatial views, morphological component analysis is deployed to the first PC of the original HSI, where the coarse, fine, high-contrast, low-contrast, horizontal and vertical features are extracted to form a matrix as another view. Then the SSMLC can be implemented.

III. Experiments and Results

A. Datasets and Parameters

In experiments, two hyperspectral image datasets, i.e., SalinasA and University of Pavia, are applied. The SalinasA dataset is acquired by the Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) sensor over the Valley of Salinas, Central Coast of California, in 1998, which includes 16 classes and 5864 labeled pixels. The University of Pavia HSI was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over Pavia, Italy. Totally, there are 9 classes and 42776 labeled samples. Considering computational cost, 1500 labeled pixels in SalinasA and 2000 labeled pixels in University of Pavia are used to validate the performance of the proposed SSMLC.

The correlation coefficient threshold in spectral partitioning is set to be {0.6, 0.7, 0.8, 0.9}. In PCA, the number of PCs is chosen as {3, 5, 10, 30, 50, 100}. After the spectral partitioning and morphological component analysis in spatial domain, obtained multi-views are usedfor MLRSSC, where the weight of each view is assumed equal, α_1 and α_2 are tuned between $[0.1, 0.9]$, and λ is chosen between $[10^{-2}, 10^{7}]$. In the SalinasA dataset, $\lambda = 10^{6}$, $\alpha_1 = 0.7$, $\alpha_2 = 0.9$ are selected for the following experiments. For the University of Pavia dataset, $\lambda = 0.1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.2$ are chosen for the following experiments.

The original SSC, single view LRSSC, and spectral-spatial sparse subspace clustering (S4C) [18] are compared with the proposed SSMLC.

B. Result and Anlysis

For the SalinasA dataset, the clustering accuracy with varied correlation coefficient threshold is shown in Fig. 1. According to the result, clustering accuracy has the maximum value 84.29% when the threshold is set to be 0.8 in spectral partitioning. There is a significant impact on spectral partitioning, where inappropriate partitioning could decrease the results more than 10%. As the threshold is chosen as 0.8, there are 29 groups after spectral partitioning, which means there are 29 views based on the spectral features in the following experiments.

Fig. 2 shows the clustering results with varied PCs in SalinasA. When PCA is applied to the original dataset alone (blue bar in Fig. 2), the accuracy is improved with the number of PCs increased to 10, then maintains stable after 10 PCs. The accuracy reaches the peak around 50 PCs, which is 85%. In this case, major information in the original HSI are contained in the first 10 PCs, and there is more noise than useful information after 50 PCs. When spectral partitions are jointly used with PCs (orange bar in Fig. 2), there is significant improvement, where the highest accuracy reaches 94% when 50 PCs view is used with 29 spectral partition views for the SSMLC algorithm.

When the morphological view is added to PCs and spectral partitions views (gray bar in Fig. 2), the highest accuracy also stays around 94%. Furthermore, the performance when using the major PCs has significant improvement, where the accuracy with 3 PCs increases around 12%, and the accuracy is stable with the number of PCs is increased. The information of spatial morphology could compensate the spectral partition information, where Fig. 2 shows the total combination improves in all other conditions. It has been shown that the PC view, spectral partition views and morphological view could provide different perspectives of observations or representations for the original data, and the combination of multiple views indeed offers the advantage for clustering.

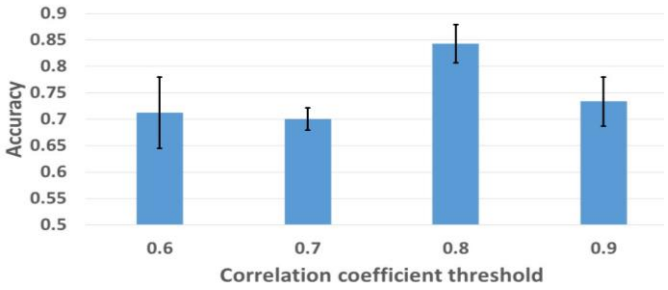

Fig. 1 clustering accuracy with different correlation coefficient threshold.

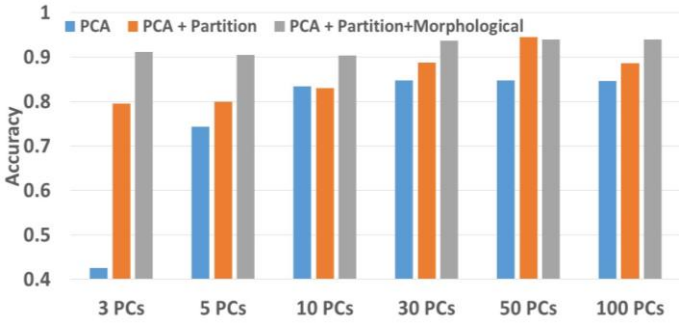

Fig. 2 The clustering accuracy with varied PCs in the SalinasA experiment.

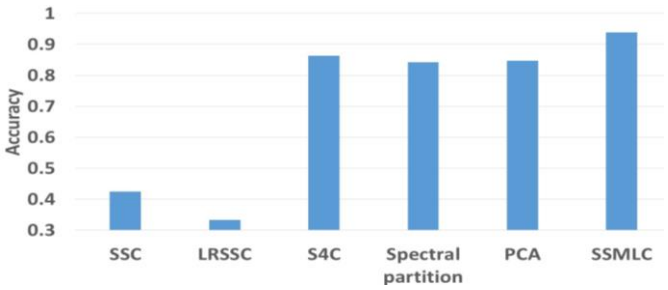

Fig. 3 The clustering accuracy with different clustering algorithms in SalinasA.

As shown in Fig. 3, several clustering algorithms are compared with the SSMLC, which include SSC, LRSSC, S4C, MLRSSC with spectral partition only, and MLRSSC with PC only. According to the results, SSMLC provides the best performance in clustering, which could yield the accuracy as high as 94%. It is truly impressive for an unsupervised approach. On the other hand, the classical SSC and LRSSC produce poor performance, which is below 50%. Spectral based MLRSSC (with spectral partition views) and the MLRSSC with the PC view have good performance, which is around 84%. But they could be improved by combination of

spectral and spatial views as in SSMLC, where the improvement for clustering accuracy is around 10%. For S4C, it has better performance than spectral based MLRSSC and PCA based MLRSSC. However, it is inferior to the proposed SSMLC for around 7%.

In the second experiment, University of Pavia dataset is investigated. The clustering results of different algorithms are shown in Fig. 4. The results further demonstrate that the proposed SSMLC generated the best performance in clustering, which is around 78%.

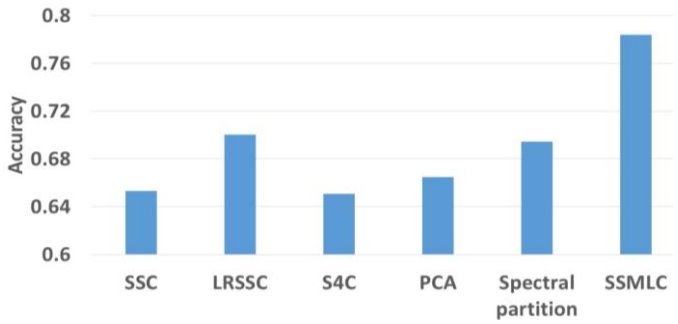

Fig. 4 The clustering accuracy with different clustering algorithms in University of Pavia.

IV. Conclusion

In this paper, the MLRSSC is extended to hyperspectral image clustering, where multiple views are extracted by spectral partitioning, morphological filtering, and PCA. The proposed SSMLC outperforms other methods of the similar type, such as SSC, LRSCC, S4C. In particular, the SSMLC offers better performance than S4C that also uses both spatial and spectral information in SSC. This means explicitly considering different types of information as multi-view in the objective function of MLRSSC can yield a more accurate affinity matrix for spectral clustering.

Similar to other spectral clustering techniques, the SSMLC is computationally expensive. We will investigate a distributed approach suitable to large-scale image clustering. We will also investigate the option to in which SSMLC will be used to learn the manifolds by learning graph affinity matrices on subset of HSI (treated as training sample) and then predicting manifold memberships on the rest (out-of-sample) pixels by using, as an example, multivariate kernel ridge regression [19].

References

[1] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and remote sensing magazine* vol. 1, pp. 6-36, Jun. 2013.

[2] M. Farid, and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on geoscience and remote sensing,* vol. 42, pp. 1778-1790, Aug. 2004.

[3] A. Sumarsono and Q. Du, "Low-rank subspace representation for estimating the number of signal subspaces in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6286-6292, Nov. 2015.

[4] W. Sun, A. Halevy, J. Benedetto, W. Czaja, C. Liu, H. Wu, B. Shi, and W. Li, "UL-Isomap based nonlinear dimensionality reduction for hyperspectral imagery classification," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 89, pp. 25-36, 2014.

[5] X. Bian, C. Chen, Y. Xu, and Q. Du, "Robust Hyperspectral Image Classification by Multi-Layer Spatial-Spectral Sparse Representations," *Remote Sensing,* vol. 8, pp. 985, 2016.

[6] E. Elhamifar, and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, pp. 2765-2781, 2013.

[7] P. Favaro, R. Vidal, A. Ravichandran, "A closed form solution to robust subspace estimation and clustering," *CVPR*, pp. 1801–1807, 2011.

[8] A. Sumarsono and Q. Du, "Low-rank subspace representation for supervised and unsupervised classification of hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, pp. 4188-4195, Sep. 2016.

[9] Y. Wang, H. Xu, and C. Leng, "Provable subspace clustering: When LRR meets SSC," In *Advances in Neural Information Processing Systems*, pp. 64-72. 2013.

[10] J. Li, X. Huang, P. Gamba, J. Bioucas-Dias, L. Zhang, J. Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing* vol, 53, pp. 1592-1606, Mar. 2015.

[11] M. Brbić, and I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognition,* vol. 73, pp. 247-258, Jan. 2018.

[12] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," *in Proc. IEEE CVPR*, pp. 1600-1607, 2012.

[13] H. Shen and J. Z. Huang, "Sparse principal component analysis via regularized low rank matrix approximation," *J. Multivariate Anal*., vol. 99, pp. 1015–1034, Jul. 2008.

[14] R. Rubinstein, R. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process*., vol. 58, pp. 1553–1564, Mar. 2010.

[15] G.Liu, Z.Lin, Y.Yu, "Robust subspace segmentation by low-rank representation," *in: Proceedingsof the 27th International Conferenceon Machine Learning (ICML-10),* pp.663–670, 2010.

[16] C. Xu, D. Tao, and C. Xu. "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634,* 2013.

[17] A. Y. Ng, M. I. Jordan, Y. Weiss, "On spectral clustering: Analysis and an algorithm," in: Advances in Neural Information Processing Systems 14, 2001, pp. 849–856.

[18] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 54, pp. 3672-3684, Jun. 2016.

[19] F. de Morsier, M. Borgeaud, V. Gass, J. P. Thiran, and D. Tuia, "Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens*., vol. 54, no. 6, pp. 3410–3420, 2016.