Multi-view Subspace Clustering for Hyperspectral Images

Shaoguang Huang and Aleksandra Pižurica
TELIN-GAIM, Ghent University, Belgium

***Abstract*— In this paper, we propose a multi-view subspace clustering model for hyperspectral remote sensing images that makes use of rich complementary information from multiple data sources (views). To capture the nonlinear data structure in each view, we introduce a novel type of spatial regularization based on hybrid hypergraphs. The constructed hybrid-hypergraph in each view consists of a series of multi-scale local hypergraphs and a spatially non-local hypergraph, which enables a comprehensive analysis of image data. Experimental results on real data demonstrate superior performance compared to the state-of-the-art.**

1 Introduction

Hyperspectral images (HSIs) acquired by airborne or satellite-borne sensors measure the objects on the Earth's surface with hundreds of spectral bands, offering this way far better discrimination between different materials than conventional multispectral images. Clustering of HSIs, which groups similar pixels in an unsupervised manner, is crucial in a number of applications in remote sensing.

Subspace clustering methods based on a self-representation model, where the input data is used as a dictionary, enjoy great success in HSI clustering. Let $\mathbf{X} \in \mathbb{R}^{B\times N}$ be the input HSI, where B denotes the number of bands and N the number of pixels. Each column of $\mathbf{X}$ represents a spectral signature in a given pixel. The sparse coding problem is defined as follows:

$$\arg\min_{\mathbf{A}} \Psi(\mathbf{X}-\mathbf{X}\mathbf{A}) + \lambda\Phi(\mathbf{A}),\ s.t.\ \mathrm{diag}(\mathbf{A}) = \mathbf{0}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the coefficient matrix of $\mathbf{X}$; $\Psi(\mathbf{X}-\mathbf{X}\mathbf{A})$ is a given loss function accounting for the data fidelity, e.g., $\|\cdot\|_F^2$ or $\|\cdot\|_1$; $\Phi(\mathbf{A})$ is a regularization term, encoding a priori knowledge about $\mathbf{A}$; λ is a parameter that balances the trade-off between the data-fit and regularization terms. The representative low-rank representation (LRR) [1] and sparse subspace clustering (SSC) [2] models utilize $\|\mathbf{A}\|_*$ and $\|\mathbf{A}\|_1$ to promote the low-rank property and sparseness of the representation matrix, respectively. Recent works improve the clustering performance by using spatial regularization, such as the centralized smoothing regularization [3], ℓ_2 norm based regularization [4] and the $\ell_{1,2}$ norm based joint representation [5, 6]. Their resulting coefficients matrix $\mathbf{A}$ is combined with its transpose to yield a symmetric similarity matrix: $\mathbf{W} = (|\mathbf{A}| + |\mathbf{A}^T|)/2$, which is then applied within the standard spectral clustering.

The main limitation of the aforementioned methods is that they are designed for processing single-source data without considering the complementary information from other data sources/features that can be useful to uncover data clusters. A recent multi-view sparse subspace clustering method for HSI [7] fails to employ the spatial information, and due to this its performance is sometimes inferior to single-view methods. In this paper, we propose a multi-view subspace clustering model that combines the information from multiple data sources and employs effectively local and non-local spatial information from each of the views.

2 Preliminaries

A hypergraph is a generalization of a normal graph where edges are replaced by hyperedges, which can connect more than two vertices simultaneously. The connected vertices correspond to entities with similar characteristics. By enabling simultaneous connections among the groups of vertices, the hypergraph encodes effectively high-order geometric data structure. We denote by $\mathcal{G}_h = (\mathbb{V}, E_h, \mathbf{W}_h)$ a hypergraph where $\mathbb{V} = \{v_i\}_{i=1}^N$ is the set of vertices corresponding to all the data points; $E_h = \{e_i\}_{i=1}^M$ is a collection of subsets of $\mathbb{V}$ and each e_i is called a hyperedge of $\mathcal{G}_h$; $\mathbf{W}_h$ is a diagonal matrix for the hyperedge weights. An incidence matrix $\mathbf{H} \in \mathbb{R}^{N\times M}$ represents the connections of vertices within each hyperedge, defined as:

$$h(v_i, e_j) = \begin{cases} 1, & if\ v_i \in e_j \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The vertex degree of each vertex $v_i \in \mathbb{V}$ and the edge degree of each hyperedge e_i are given by

$$d(v_i) = \sum_{e_j \in E_h} W_{h_{jj}} h(v_i, e_j) \tag{3}$$

$$d(e_j) = \sum_{v_i \in \mathbb{V}} h(v_i, e_j). \tag{4}$$

Denote by $\mathbf{D}_v$ and $\mathbf{D}_e$ two diagonal matrices with $\mathbf{D}_{v_{ii}} = d(v_i)$ and $\mathbf{D}_{e_{ii}} = d(e_i)$. We formulate a manifold constraint based on the hypergraph by $\Gamma(\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{L}\mathbf{A}^T)$, where $\mathbf{L} = \mathbf{D}_v - \mathbf{H}\mathbf{W}_h\mathbf{D}_e^{-1}\mathbf{H}^T$ is the Laplacian matrix of $\mathcal{G}_h$.

3 Proposed Method

Let $\{\mathbf{X}^t \in \mathbb{R}^{B_t\times N}\}_{t=1}^T$ denote the multi-view data, where B_t is the dimension of the t-th data and T is the number of data sources. The proposed multi-view subspace clustering model is formulated as follows:

$$\begin{aligned} \min\ & \sum_{t=1}^{T}(\|\mathbf{X}^t - \mathbf{X}^t\mathbf{A}^t\|_F^2 + \lambda_1\Theta(\mathbf{A}^t) \\ & \qquad + \lambda_2\|\mathbf{E}^t\|_1) + \lambda_3\|\mathbf{Z}\|_* \\ s.t.\ & \mathbf{A}^t = \mathbf{Z} + \mathbf{E}^t\ \ (\forall t = 1, 2, ..., T) \end{aligned} \tag{5}$$

where λ_1, λ_2 and λ_3 are positive numbers, $\mathbf{A}^t$ is the coefficient matrix of $\mathbf{X}^t$ and $\Theta(\mathbf{A}^t) = \Gamma_l(\mathbf{A}^t) + \Gamma_n(\mathbf{A}^t)$ is a hybrid-hypergraph-based regularization. $\Gamma_l(\mathbf{A}^t)$ and $\Gamma_n(\mathbf{A}^t)$ are defined respectively by $\sum_i^p \mathrm{tr}(\mathbf{A}^t\mathbf{L}_{l_i}^t{\mathbf{A}^t}^T)$ and $\mathrm{tr}(\mathbf{A}^t\mathbf{L}_n^t{\mathbf{A}^t}^T)$, where $\mathbf{L}_{l_i}^t$ is the i-th Laplacian matrix of multi-scale local hypergraphs, p is the number of scales, and $\mathbf{L}_n^t$ is the Laplacian matrix of a spatially non-local hypergraph in t-th data source.

The global low-rank matrix $\mathbf{Z}$ in (5) is shared by all the data sources and indicates the common underlying low-rank structure in the lower-dimensional subspaces, while the additive sparse matrices $\mathbf{E}^t$ represent the deviations from the consensus matrix. The regularization $\Theta(\mathbf{A}^t)$ is utilized to model the important local and nonlocal spatial information in each data source, which facilitates the uncovery of the intrinsic data cluster structure. Next, we explain how we construct the hybrid-graph for each view. For simplicity, we take HSI as an example by removing the index t. Other available imaging data sources can be incorporated in similar way.

To exploit the local and non-local spatial information of HSIs, we build the hybrid-hygergraph composed of two types of hypergraphs: multi-scale local hypergraphs $\mathcal{G}_l^i$ $(i = 1, ..., p)$ and a nonlocal hypergraph $\mathcal{G}_n$. Specifically, we first utilize a super-pixel segmentation approach [8] to segment HSIs into non-overlapping super-pixels under different segmentation granularities, and then based on the segmentation results we build the multi-scale local hypergraphs. With varying number of super-pixels $n_i(i = 1, ..., p)$, we obtain super-pixel segmentation maps at different scales. To decide the values of n_i, we adopt a simple strategy by setting $n_i = 2^{i-1} \cdot n$ and $p = 4$, where n is the smallest value of n_i. We calculate n empirically as $n = \lfloor \frac{1}{2\sqrt{N}} \|\nabla \mathbf{f}_s\|_1 \rfloor$. $\nabla \mathbf{f}_s \in \mathbb{R}^{1\times N}$ is a vector of thresholded gradient components

$$\nabla \mathbf{f}_{s_i} = \begin{cases} 1, & if\ \nabla \mathbf{f}_i > \delta \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where $\nabla \mathbf{f}$ is the gradient of the first principle component of $\mathbf{X}$, and the threshold δ is defined as $\delta = \frac{1}{N}\sum_i \nabla f_i$. The threshold excludes the less relevant edges from $\nabla \mathbf{f}$ and leaves only the significant ones for further consideration.

By setting the number of super-pixels as $\{n_i\}_{i=1}^4$, we obtain a set of super-pixels segmentation maps $\{\mathbf{F}_i \in \mathbb{R}^N\}_{i=1}^4$, with each $\mathbf{F}_i = \{y_1^i, y_2^i, ..., y_N^i\}$, where y_j^i is an integer between 1 and n_i, indicating the label of pixel $\mathbf{x}_j$ in i-th level segmentation. We view each pixel as a vertex and each super-pixel as a hyperedge, and we calculate the resulting incidence matrix $\mathbf{H}^i$ for the multi-scale local hypergraph $\mathcal{G}_l^i$ as follows:

$$h^i(v_j, e_k^i) = \begin{cases} 1, & if\ v_j \in e_k^i \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $e_k^i = \{v_l\}_{l\in\{j|y_j^i=k\}}$ is the k-th hyperedge in the i-th hypergraph. We calculate the diagonal hyperedge weights matrices as $W_{h_{jj}}^i = \sum_{\mathbf{x}_k,\mathbf{x}_l\in e_j^i} \exp(-\|\mathbf{x}_k - \mathbf{x}_l\|^2/\sigma_1^2)$. The corresponding matrices $\mathbf{D}_v^i$ and $\mathbf{D}_e^i$ for vertex degree and edge degree are obtained similarly by (3) and (4). We further calculate the Laplacian matrices of local hypergraphs by $\mathbf{L}_{l_i} = \mathbf{D}_v^i - \mathbf{H}^i\mathbf{W}_h^i(\mathbf{D}_e^i)^{-1}\mathbf{H}^{i^T}$.

For the non-local hypergraph $\mathcal{G}_n$, we first extract centralized patches for all the pixels by using a square window $p \times p$. Let $\mathbf{X}_{p_i} = [\mathbf{x}_i, \mathbf{x}_{i_1}, ..., \mathbf{x}_{i_{R-1}}] \in \mathbb{R}^{B\times R}$ denote the hyperspectral patches for the central pixel $\mathbf{x}_i$, where $R = p^2$ is the number of pixels in a patch and $\{\mathbf{x}_{i_j}\}_{j=1}^{R-1}$ are the corresponding neighbours of $\mathbf{x}_i$. Then, we exploit each pixel and its K closest neighbours to construct the hyperedges $E_h^n = \{e_i^n\}_{i=1}^N$ based on the patch-wise similarity, measured by $s(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{X}_{p_i} - \mathbf{X}_{p_j}\|_F^2/\sigma_2^2)$. Let $\mathcal{N}(\mathbf{x}_i)$ represent the index set of the neighbours of pixel $\mathbf{x}_i$. We derive the incidence matrix $\mathbf{H}^n$ for the spatial-wise nonlocal hypergraph $\mathcal{G}_n$ as

$$h^n(v_i, e_j^n) = \begin{cases} 1, & if\ v_i \in e_j^n \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $e_j^n = \{v_i\}_{i\in\mathcal{N}(\mathbf{x}_j)}$ is the j-th hyperedge. The values of hyperedge weights $\mathbf{W}_h^n$ are calculated by $W_{h_{jj}}^n = \sum_{\mathbf{x}_k,\mathbf{x}_l\in e_j^n} s(\mathbf{x}_k, \mathbf{x}_l)$. We calculate vertex degree and edge degree matrices $\mathbf{D}_v^n$ and $\mathbf{D}_e^n$ similarly using (3) and (4), and we obtain the Laplacian matrix of non-local hypergraph as $\mathbf{L}_n = \mathbf{D}_v^n - \mathbf{H}^n\mathbf{W}_h^n(\mathbf{D}_e^n)^{-1}\mathbf{H}^{n^T}$.

We solve the optimization problem (5) using the alternating direction method of multipliers. Once obtaining the global consensus matrix $\mathbf{Z}$, we construct the similarity matrix as $\mathbf{W} = (|\mathbf{Z}| + |\mathbf{Z}^T|)/2$, which is further applied in the standard spectral clustering [9] to obtain the clustering result.

Table 1: Clustering results in *Indian Pines*

Methods	k-means	SSC	JSSC	SSMLC	HMSC
OA(%)	66.91	68.00	89.05	79.23	94.28
κ	0.48	0.53	0.85	0.70	0.92
NMI	0.51	0.45	0.73	0.53	0.83

Table 2: Clustering results in *Houston*

Methods	k-means	SSC	JSSC	SSMLC	HMSC
OA(%)	74.97	73.18	79.54	74.05	82.18
κ	0.69	0.65	0.75	0.67	0.78
NMI	0.73	0.79	0.78	0.80	0.83

4 Results and Discussion

We test our hybrid-hypergraph based multi-view subspace clustering (HMSC) on two real remote sensing data sets: *Indian Pines* and *Houston*. The *Indian Pines* image is of size $85 \times 70 \times 200$, and contains four classes. We employ its extended multiattribute profiles (EMAPs) spatial feature as a second data source. For more details of EMAPs, we refer to [10]. *Houston* contains registered HSI and pseudowaveform LiDAR. The hyperspectral image used in our tests is of size $130 \times 130 \times 144$, and has seven classes. To increase the discriminative ability of LiDAR, we follow [11] and employ the EMAPs spatial feature of LiDAR as the second data source.

We compare the performance of the proposed method with three single-view clustering methods k-means [12], SSC [2] and joint SSC (JSSC) [6], and a multi-view clustering method SSMLC [7]. The clustering results are reported in Tables 1 and 2 with three quantitative evaluation metrics: overall accuracy (OA), kappa (κ) and Normalized Mutual Information (NMI). We refer to [13] for the details of clustering evaluation. For single-view clustering methods, we report the clustering results with the data source that yields the highest OA. In *Indian Pines*, we find that all the single-view clustering methods produce a higher accuracy with EMAPs spatial feature. On the contrary, in *Houston*, HSI yields better performance. This demonstrates the varying abilities of data sources in uncovering cluster structure. The results in Table 1 and 2 show that our HMSC achieves the best result in terms of OA, κ and NMI. The multi-view SSMLC method performs better than the single-view SSC model, but worse than JSSC. This can be attributed to the fact that JSSC takes local spatial information into account, and SSMLC does not. Our model exploits the rich complementary information from multi-view data and employs jointly the local and nonlocal spatial information in each view, facilitating thereby the effective uncovery of intrinsic data cluster structure.

References

[1] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, 2012.

[2] 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.

[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, 2016.

[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, 2016.

[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] S. Huang, H. Zhang, and A. Pižurica, "Semisupervised sparse subspace clustering method with a joint sparsity constraint for hyperspectral remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 989–999, 2019.

[7] L. Tian, Q. Du, I. Kopriva, and N. Younan, "Spatial-spectral based multi-view low-rank sparse subspace clustering for hyperspectral imagery," in *Proc. IEEE IGARSS*, 2018, pp. 8488–8491.

[8] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2097–2104.

[9] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[10] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, 2010.

[11] Y. Zhang and S. Prasad, "Multisource geospatial data fusion via local joint sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3265–3276, 2016.

[12] S. Lloyd, "Least squares quantization in pcm," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[13] S. Huang, H. Zhang, Q. Du and A. Pižurica, "Sketch-Based Subspace Clustering of Hyperspectral Images," *Remote Sensing*, vol. 12, no. 5, 2020.