Hybrid-hypergraph Regularized Multi-view Subspace Clustering for Hyperspectral Images

Shaoguang Huang, Member, IEEE, Hongyan Zhang, Senior Member, IEEE, and Aleksandra Pižurica, Senior Member, IEEE

Abstract—Clustering algorithms play an essential and unique role in classification tasks, especially when annotated data are unavailable or very scarce. Current clustering approaches in remote sensing are mostly designed for a single data source, such as hyperspectral image (HSI), while nowadays multi-sensor data are being routinely acquired. In this paper, we propose a multiview subspace clustering model, which exploits effectively the rich information from multiple features extracted either from a single data source (HSI) or from multiple sources, that we call generically multi-views of the same scene. An important novelty of our approach is that it integrates local and nonlocal spatial information from each view in a unified framework. Our model learns a common intrinsic cluster structure from viewspecific subspace representations by a new decomposition-based scheme. In addition, we develop innovative manifold-based spatial regularization as a hybrid hypergraph, which merges local and non-local spatial context and improves thereby the learning of view-specific structures. We develop an efficient algorithm to solve the resulting optimization problem. Extensive experiments on real datasets demonstrate the superior clustering performance over the state-of-the-art.

Index Terms-Hyperspectral images, remote sensing, subspace clustering, multi-view clustering.

I. INTRODUCTION

YPERSPECTRAL imaging systems measure the light reflected from object reflected from objects in hundreds of spectral bands, covering the spectral range from visible to near-infrared. This wealth of spectral information enables far better discrimination between diverse materials compared to traditional color and multispectral images. Consequently, hyperspectral images (HSIs) find numerous applications in remote sensing, in the domains such as precision agriculture [1, 2], defense and security [3], geology and mineralogy [4] and environmental monitoring [5, 6]. In all these applications, image classification, as a fundamental step in data preprocessing, provides a basis for the automatic HSI analysis and scene interpretation.

Over the last two decades, a number of supervised classification methods for HSIs have been proposed [7–9]. Some of the most prominent approaches are based on random

H. Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zhanghongyan@whu.edu.cn).

forests [10], support vector machine [11], sparse representation classification [12] and the recently arisen deep neural networks [13]. All these rely on labeled training samples to optimize the delicately designed classifiers. Since data labeling is typically labor intensive and time-consuming, labeled data required for training the classifiers are often scarce, posing serious limitations for the supervised classification methods [14]. In contrast, clustering, as an unsupervised approach, aims to discriminate data points belonging to different clusters without using any labeled data. Thus, clustering enables automatic data processing and interpretation in cases where supervised classification is infeasible. This is especially of interest in dynamic scenarios like monitoring forest fires and disaster damages where clustering plays a unique role [15].

Common clustering approaches include hierachical clustering [16, 17], centroid-based [18-22], density-based [23], biological clustering [24] and spectral-based methods [25-29]. Spectral-based clustering methods, which are of special interest here, consist of two steps: construction of a similarity matrix and spectral clustering. A large body of the literature has focused on building a desirable similarity matrix. Representative methods are low-rank representation (LRR) model [27] and sparse subspace clustering (SSC) model [28]. While these approaches achieved great success in computer vision, their direct application in HSI clustering often yields unsatisfactory results, due to various reasons, including noise, within-class spectral variability and complex data structure [30, 31].

Various extensions of subspace clustering have been developed to alleviate these problems. Incorporating spatial information in the SSC model by using smoothing strategies in a local square window as proposed in [30] proved to be effective. Several follow-up works [32–35] exploit the spatial information using other types of spatial regularization, such as ℓ_2 norm-based smoothing term [32], joint representation [33] and total variation [34, 35]. An alternative to integrating spatial regularizations into the learning model is post-processing of the representation coefficients as in [36, 37]. Other representative approaches [38-40] build an anchor-based graph with a few initially selected samples with the aim to reduce the overall computational complexity of SSC. Generalizations to semi-supervised clustering models include [41, 42] and parameter-free multi-objective SSC was proposed in [43].

While achieving improved clustering performance, the above mentioned SSC-based clustering methods still face two crucial limitations. Firstly, they incorporate only local spatial content from windows of fixed size at a single particular scale.

This work was supported in part by the Flanders AI Research Programme under the grant 174B09119 and in part by the National Natural Science Foundation of China under the grants 61871298 and 42071322. (Corresponding author: Hongyan Zhang.)

S. Huang and A. Pižurica are with the Department of Telecommunications and Information Processing, TELIN-GAIM, Ghent University, 9000 Ghent, Belgium (e-mail: Shaoguang.Huang@ugent.be; Aleksandra.Pizurica@ugent.be).

However, it is hard to determine the optimal window size due to the varying spatial resolutions and different scenes in HSIs. Also, the experience from other image processing problems, such as denoising [44] teaches us that the use of nonlocal image similarities can result in huge improvements compared to using the local correlations only. Secondly, all of the methods mentioned above are designed for the clustering task of single-source data. Nowadays, multi-source data in remote sensing such as hyperspectral images and Light Detection and Ranging (LiDAR) are routinely acquired and making use of their complementary information can help in discriminating better between data points from different classes. Moreover, many spatial-spectral features that are extracted from HSIs, such as morphological profiles, Gabor features and local binary patterns, proved to be more effective than the original spectral data in the analysing of HSIs [45-47]. Here, we refer to the different data sources such as HSI and LiDAR and features extracted either from single-source or multi-source data as different views of the same scene. Recently, the authors in [48] reported the first attempt in multi-view HSI clustering. However, this method is sensitive to noise and outliers, and its high overall computational complexity poses limitations to the number of data sources. While multi-view clustering methods [49-52] demonstrated success in applications with facial images, online documents and hand written text, none of them was applied in the clustering of HSIs. Our experiments show that directly applying such methods in the clustering of HSIs often yields unsatisfactory performance, which can be mainly attributed to spectral noise and spectral variability.

We propose a novel multi-view subspace clustering method for HSI, which effectively utilizes the rich information from different views and integrates local and nonlocal spatial information. An important novelty is a decomposition-based learning framework for multi-view data in the low-dimensional subspaces. Specifically, the subspace representations obtained from different views are decomposed as a combination of a global low-rank consensus matrix and view-specific sparse matrices. This decomposition strategy is based on the following two important observations: 1) the view-specific representations derived from a self-representation model often violate from the ideal block-diagonal structure due to the effect of noise and large within-cluster spectral variability; 2) the ideal block-diagonal matrix presents an intrinsic lowrank property [27], which should be shared among the viewspecific representations. Thus, we use the global low-rank matrix to model the common underlying low-rank structure shared by all the views, and the sparse matrices to account for the view-specific deviations from this common structure. We define a specific regularizer as a manifold constraint based on a hybrid hypergraph that incorporates the spatial content in each view. In particular, we build the hybrid hypergraph from a series of multi-scale local hypergraphs and a spatially nonlocal hypergraph, which encode the local and nonlocal spatial information, respectively. The local hypergraphs are constructed by applying a sequence of super-pixel segmentations at different scales, which are automatically determined by the gradient information of the input image. The nonlocal hypergraph is constructed based on the patch-wise similarity. We develop an efficient optimization algorithm to solve the resulting model. Experimental results on real data sets confirm the efficacy of our method in comparison with the current state-of-the-art in the field.

The rest of this paper is organized as follows. Section II briefly reviews single-view subspace clustering methods for hyperspectral image, basic concepts behind hypergraphs and super-pixel segmentation. Section III introduces the proposed multi-view subspace clustering method and develops an efficient algorithm to solve the resulting optimization problem. Experiments and analysis on the benchmark datasets are conducted in Section IV. Section V concludes the paper.

II. PRIOR WORK AND PRELIMINARIES

The following notation will be used throughout the paper. $\|\mathbf{A}\|_1 = \sum_i \sum_j |A_{ij}|$ is the ℓ_1 norm of matrix \mathbf{A} , where A_{ij} denotes the element in *i*-th row and *j*-th column of \mathbf{A} . $\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j A_{ij}^2}$ is the Frobenius norm and $\|\mathbf{A}\|_{1,2} = \sum_i \sqrt{\sum_j A_{ij}^2}$ is the $\ell_{1,2}$ norm of \mathbf{A} . $|\mathbf{A}|$ is a matrix with each entry being the absolute value of the corresponding entry in \mathbf{A} . The nuclear norm of \mathbf{A} , $\|\mathbf{A}\|_{*}$, is the sum of singular values of matrix \mathbf{A} . We denote by diag(\mathbf{A}) a vector with its *i*-th element being A_{ii} and Diag(\mathbf{a}) a diagonal matrix with its diagonal elements being a_i . $\mathbf{1}$ is an all-one vector and $\langle \mathbf{A}, \mathbf{B} \rangle = \text{Tr}(\mathbf{A}^T \mathbf{B}) = \sum_i \sum_j A_{ij} B_{ij}$ is the inner product between \mathbf{A} and \mathbf{B} , where $\text{Tr}(\mathbf{Z})$ is the trace of a real square matrix \mathbf{Z} , i.e., $\text{Tr}(\mathbf{Z}) = \sum_i Z_{ii}$.

A. Single-view Subspace Clustering Model for HSIs

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 **X** represents a spectral signature in a given pixel. The self-representation sparse coding problem is defined as follows:

$$\underset{\mathbf{A}}{\arg\min} \Psi(\mathbf{X} - \mathbf{X}\mathbf{A}) + \lambda \Phi(\mathbf{A}), \ s.t. \ \operatorname{diag}(\mathbf{A}) = \mathbf{0}, \quad (1)$$

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 LRR [27] and SSC [28] models utilize $\|\mathbf{A}\|_*$ and $\|\mathbf{A}\|_1$ to promote the low-rank property and sparseness of the representation matrix, respectively. Recent works improve clustering performance by using spatial regularization, such as the centralized smoothing regularization in [30], ℓ_2 norm based regularization in [32] and the $\ell_{1,2}$ norm based joint representation [33].

The resulting coefficients matrix **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. Specifically, the *c* eigenvectors $\{\mathbf{v}_k\}_{k=1}^c$ of the Laplacian matrix $\mathbf{L} = \text{Diag}(\mathbf{W1}) - \mathbf{W}$ corresponding to its *c* smallest eigenvalues are first calculated and then the clustering results are obtained by applying the k-means algorithm to the matrix $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_c]$, where *c* is the number of clusters.



Fig. 1. An example of hypergraph (a) and its corresponding incidence matrix **H** (b). The hypergraph includes eight vertices $\{v_i\}_{i=1}^8$ marked by blue dots and three hyperedges $\{e_j\}_{j=1}^3$ annotated by three colored regions.

B. Hypergraph Preliminaries

A hypergraph is a generalization of an ordinary graph, where an edge, called hyperedge, can connect more than two vertices. This offers more flexibility in modelling complex topological structures of high-dimensional data [53]. 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

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

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

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

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

Fig. 1 (a) shows an example of a hypergraph with eight vertices and three hyperedges. For instance, the hyperedge e_1 connects a group of three vertices $\{v_1, v_2, v_3\}$ and e_2 connects four vertices v_2, v_4, v_5 and v_6 . The corresponding incidence matrix in Fig. 1 (b) represents compactly the connectivity structure.

C. Entropy Rate Segmentation

Entropy rate segmentation (ERS) method [54] is a graphbased clustering algorithm, which segments effectively an image into non-overlapping super-pixels, each of which consists of highly similar pixels. Given an ordinary graph G = (V, E), where V is the vertex set corresponding to the pixels of the image and E is the edge set representing the pairwise similarities between neighbouring pixels, ERS aims to segment the graph into connected sub-graphs (each corresponding to a superpixel) by selecting a subset of edges $A \subseteq E$. Mathematically, ERS solves the following problem:

$$\underset{A}{\operatorname{arg\,min}} \ \mathcal{H}(A) + \lambda \mathcal{B}(A) \ s.t. \ A \subseteq E \ \text{and} \ N_A \ge n \quad (5)$$

where n is a predefined number of super-pixels; N_A is the number of connected components in the graph; $\mathcal{H}(A)$ is the entropy rate of the random walk on G = (V, A); $\mathcal{B}(A)$ is a balancing term to encourage clusters with similar size and λ is a parameter to control the balance between the two terms. We refer to [54] for details. Note that λ can be estimated adaptively by the value of n, therefore the parameter that needs to be specified by user in ERS is n. Incorporating the two terms in the objective function, ERS is able to generate compact, homogeneous and balanced super-pixels.

III. HYBRID-HYPERGRAPH REGULARIZED MULTI-VIEW SUBSPACE CLUSTERING FOR HSIS

Here we introduce our hybrid-hypergraph regularized multiview subspace clustering (HMSC) model for hyperspectral images. We first describe our multi-view clustering model, and then present a new hybrid-hypergraph-based regularization. Finally, an efficient algorithm is proposed to solve the resulting optimization problem.

A. Multi-view Subspace Clustering

Multimodal data are commonly available now, offering much richer information, better discrimination among different classes of interest and better robustness to various degradations and missing data. For example, while HSIs enable to discriminate between different materials, LiDAR data provides a complementary information about the altitude of the imaged objects. Apart from the data acquired by different sensors, many spatial features that are dedicatedly extracted from HSIs, such as Gabor features [11, 55], morphological features [56] and local binary patterns [46], are proved to be effective in improving the classification performance of HSI. Combining such complementary sources of information enables better clustering performance than based on a single-view data alone. Here we propose an effective multi-view subspace clustering approach with the aim to enhance the clustering performance by incorporating data from multiple sources.

Let $\{\mathbf{X}^t \in \mathbb{R}^{B_t \times N}\}_{t=1}^T$ denote the multi-view data, where B_t is the dimensionality of the *t*-th data source and *T* is the number of data sources. According to (1), the data points in each view can be represented by a linear combination of others, i.e., $\mathbf{X}^t \approx \mathbf{X}^t \mathbf{A}^t$, where \mathbf{A}^t is the coefficient matrix of \mathbf{X}^t . Ideally, the representation matrices \mathbf{A}^t show a block-diagonal structure [28], where the coefficients corresponding to the data points from distinct clusters are zero. However, this can be rarely true due to the effect of noise and large spectral variability within class [42]. Observe that a block-diagonal matrix presents an intrinsic low-rank property [27]. To capture the underlying common data structure, we propose a low-rank decomposition based scheme, which decomposes the view-specific matrix \mathbf{A}^t as a global low-rank matrix $\mathbf{Z} \in \mathbb{R}^{N \times N}$ and a sparse matrix $\mathbf{E}^t \in \mathbb{R}^{N \times N}$. \mathbf{Z} is shared by all the subspace



Fig. 2. An example by using HSI for the construction of hybrid-hypergraph, including multi-scale local hypergraphs (top) and spatial-wise nonlocal hypergraph (bottom).

representation matrices \mathbf{A}^t to model the consensus between different data sources, while \mathbf{E}^t models the corresponding errors. The proposed multi-view subspace clustering model is formulated as follows:

$$\min \sum_{t=1}^{T} (\|\mathbf{X}^{t} - \mathbf{X}^{t} \mathbf{A}^{t}\|_{F}^{2} + \lambda_{1} \Theta(\mathbf{A}^{t}) + \lambda_{2} \|\mathbf{E}^{t}\|_{1}) + \lambda_{3} \|\mathbf{Z}\|_{*}$$

s.t. $\mathbf{A}^{t} = \mathbf{Z} + \mathbf{E}^{t} \quad (\forall t = 1, 2, ..., T)$ (6)

where λ_1 , λ_2 and λ_3 are positive numbers, and $\Theta(\mathbf{A}^t)$ is adopted regularization to model the important local and nonlocal spatial information in t-th data source. We elaborate on $\Theta(\mathbf{A}^t)$ in the following part. The global consensus low-rank matrix Z in (6) links all the representation matrices A^{t} and reveals the underlying common data structure of all the data sources in the low-dimensional subspaces. The computational complexity of solving a low-rank regularized problem is $\mathcal{O}(N^3)$ [57]. Compared with the multi-view HSI clustering method [48] which integrates a low-rank regularization for each \mathbf{A}^t , our model contains only one low-rank related constraint, achieving thereby a lower computational complexity. Moreover, our adopted spatial regularization $\Theta(\mathbf{A}^t)$ integrates both local and nonlocal topological information of each data source in the clustering model, facilitating the learning model to uncover a more precise cluster structure than [48].

Once obtaining the global consensus matrix \mathbf{Z} , we construct a similarity matrix by $\mathbf{W} = (|\mathbf{Z}| + |\mathbf{Z}^T|)/2$, which is further applied in the standard spectral clustering [58] to obtain the clustering result.

B. Hybrid-hypergraph Regularization

The current subspace clustering approaches for hyperspectral images typically consider only the local spatial information in the learning model, extracted from fixed-size windows at a particular scale. The main idea behind our approach is to exploit 1) multiscale spatial information (instead of setting fixed window sizes) and 2) non-local similarities of image features (instead of relying on local spatial correlations alone). We shall develop a unified framework that merges these components elegantly within a hybrid hypergraph. Fig. 2 illustrates an example by using HSI for the construction of this hybrid hypergraph. In the following, we suppress for compactness the index t that refers to a particular data source.

1) Multi-scale local hypergraph: Although spatial context modelling has been a very active area in image processing in general, the current clustering models for hyperspectral images typically ignore the spatial context or incorporate relatively simple features like pairwise correlations extracted from fixed-size windows [30, 32–34, 37, 59]. Here we build instead a multiscale spatial context model, by applying a sequence of super-pixel segmentations at different granularity levels. Each of these segmentation levels is encoded in a local hypergraph at the corresponding scale. The multiscale spatial content is expressed through a sequence of local hypergraphs as illustrated in the top of Fig. 2.

Each super-pixel can be viewed as a group of locally confined pixels that share similar characteristics [54]. We view the pixels within one super-pixel as a hyperedge of the hypergraph at the corresponding scale. Directly applying a super-pixel segmentation method as [54] in HSI is infeasible as hyperspectral images contain hundreds of spectral bands. As most of these bands are highly correlated, the first principal component (PC) depicts reasonably well the dominant spatial structures in the whole hyperspectral data cube and is employed as the input image for super-pixel segmentation. Determining the number of super-pixels n (which translates directly to the corresponding granularity of the segmentation) involves always a trade-off. A small n leads to large super-pixels, that may include pixels from multiple classes. Conversely, a too large nresults in tiny super-pixels with a cumbersome representation and loss of important local correlations. A proper n needs to be determined for each image at hand, depending on its resolution. We avoid these problems by introducing a sequence of segmentation maps instead of choosing a particular n.

Let $\{\mathbf{F}_i\}_{i=1}^p$ denote the sequence of segmentation maps at p scales, where $\mathbf{F}_i = [l_{i1}, l_{i2}, ..., l_{iN}]$ and $l_{ij} \in \{1, 2, ..., n_i\}$ is the label of the pixel j in the *i*-th segmentation map obtained by using ERS [54] in the first PC of HSI. $n_i = 2^{i-1}n$ is the number of super-pixels at scale i. We determine n based on the gradient information. The rationale is that thresholded gradient magnitudes indicate important image discontinuities, and the amount of these discontinuities is related to the number of super-pixels. Denote by $\mathbf{f} \in \mathbb{R}^{1 \times N}$ the first PC of \mathbf{X} , and its gradient $\nabla \mathbf{f}$ by

$$\nabla \mathbf{f} = |\mathbf{H}_x \mathbf{f}^T| + |\mathbf{H}_y \mathbf{f}^T|, \tag{7}$$

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. It is reasonable to segment image into more super-pixels when more spatial edges are present. Motivated by this, we define

$$n = \lfloor \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} (\nabla f_i - \delta)_+ \rfloor$$
(8)

where $|\cdot|$ denotes the rounding floor operation and

$$(x)_{+} = \begin{cases} 1, & \text{if } x > 0\\ 0, & \text{otherwise} \end{cases}$$
(9)

 δ is a threshold defined as $\delta = \frac{1}{N} \sum_{i} \nabla f_{i}$. The threshold excludes the less relevant edges from $\nabla \mathbf{f}$ and only takes the significant ones into consideration.

With equation (8), the segmentation now is adaptive to the image size, spatial content and image resolution. By setting the number of super-pixels as $\{n_i\}_{i=1}^p$, we obtain a set of super-pixels segmentation maps $\{\mathbf{F}_i\}_{i=1}^p$. We view each pixel as a vertex and each super-pixel as a hyperedge and correspondingly, we define the incidence matrix \mathbf{H}^i for the multi-scale local hypergraph \mathcal{G}_l^i as

$$H_{jk}^{i} = h^{i}(v_{j}, e_{k}^{i}) = \begin{cases} 1, & \text{if } v_{j} \in e_{k}^{i} \\ 0, & \text{otherwise,} \end{cases}$$
(10)

where $e_k^i = \{v_l\}_{l \in \{j | l_{ij} = k\}}$ is the k-th hyperedge in the *i*-th hypergraph. The diagonal elements of the matrix \mathbf{W}_h^i for the hyperedge weights are given by

$$W_{h_{jj}}^{i} = \sum_{\mathbf{x}_k, \mathbf{x}_l \in e_j^i} \exp(-\frac{\|\mathbf{x}_k - \mathbf{x}_l\|^2}{\sigma_1^2}), \quad (11)$$

where $\sigma_1 = \frac{1}{N^2} \sum_{k,l} \|\mathbf{x}_k - \mathbf{x}_l\|_2$ is the mean pairwise distance in the given dataset. For the sake of further analysis, it will be convenient to express the vertex degrees and the edge degrees as diagonal matrices. We denote by $\mathbf{D}_v^i \in \mathbb{R}^{N \times N}$ the diagonal matrix that contains the degrees of vertices at scale i and define the entries $D_{v_{ij}}^i$ as

$$D_{v_{jj}}^{i} = d^{i}(v_{j}) = \sum_{e_{k} \in E_{h}^{i}} W_{h_{kk}}^{i} h^{i}(v_{j}, e_{k}^{i}),$$
(12)

where E_h^i is the collection of hyperedges in \mathcal{G}_l^i . Let a diagonal matrix $\mathbf{D}_e^i \in \mathbb{R}^{n_i \times n_i}$ represent the edge degrees at scale *i* with diagonal elements $D_{e_{ii}}^i$ defined as

$$D_{e_{jj}}^{i} = d^{i}(e_{j}^{i}) = \sum_{v_{i} \in \mathbb{V}} h^{i}(v_{i}, e_{j}^{i}).$$
(13)

The above defined vertex degree \mathbf{D}_v^i and edge degree \mathbf{D}_e^i matrices will be employed to formulate compactly the regularization function.

2) Nonlocal hypergraph: Images of natural scenes, including hyperspectral images in remote sensing, exhibit nonlocal self-similarities: similar patches occur at different locations in the image [60]. To capture this nonlocal spatial information, we design an additional hypergraph \mathcal{G}_n that we shall refer to as nonlocal hypergraph. The whole procedure is depicted in the bottom of Fig. 2. Firstly, we extract centralized patches for all the pixels by using a square window $w \times w$. Let \mathcal{X}_i denote the data cube centered at spatial location *i*, extending over the spatial window $w \times w$ and covering all the spectral bands. We construct a hyperedge for each pixel and its *K* closest neighbours in terms of the patch-wise similarity, measured as follows

$$s(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{\|\mathcal{X}_i - \mathcal{X}_j\|_F^2}{\sigma_2^2}),$$
(14)

where $\sigma_2 = \frac{1}{N^2} \sum_{i,j} \|\mathcal{X}_i - \mathcal{X}_j\|_F$ is the mean distance of all the patches in HSI. Let $\mathcal{N}(\mathbf{x}_i)$ represent the set of indices of the nearest neighbours of \mathbf{x}_i in terms of the similarity measure defined above. We define the incidence matrix \mathbf{H}^n for the nonlocal hypergraph \mathcal{G}_n as

$$H_{ij}^n = h^n(v_i, e_j^n) = \begin{cases} 1, & \text{if } v_i \in e_j^n \\ 0, & \text{otherwise,} \end{cases}$$
(15)

where $e_j^n = \{v_i\}_{i \in \mathcal{N}(\mathbf{x}_j)\}}$ is the *j*-th hyperedge. Let \mathbf{W}_h^n be a diagonal matrix containing the hyperedge weights. The diagonal elements are given by

$$W_{h_{jj}}^{n} = \sum_{\mathbf{x}_{k}, \mathbf{x}_{l} \in e_{j}^{n}} s(\mathbf{x}_{k}, \mathbf{x}_{l}).$$
(16)

The vertex degree and the edge degree in the nonlocal hypergraph \mathcal{G}_n are denoted by diagonal matrices \mathbf{D}_v^n and \mathbf{D}_e^n , respectively. We calculate the entries $D_{v_{ij}}^n$ and $D_{e_{ij}}^n$ as

$$D_{v_{jj}}^{n} = d^{n}(v_{j}) = \sum_{e_{k} \in E_{h}^{n}} W_{h_{kk}}^{n} h^{n}(v_{j}, e_{k}^{n}), \qquad (17)$$

$$D_{e_{jj}}^{n} = d^{n}(e_{j}^{n}) = \sum_{v_{i} \in \mathbb{V}} h^{n}(v_{i}, e_{j}^{n}).$$
(18)

3) Hybrid-hypergraph regularization: The aforementioned two types of hypergraphs capture the important spatial information of hyperspectral data from local and nonlocal perspectives. We derive the hybrid hypergraph \mathcal{G}_h directly from the combination of the multi-scale local hypergraphs \mathcal{G}_l^i and

the spatial-wise nonlocal hypergraph \mathcal{G}_n , and formulate the proposed hybrid-hypergraph-based regularization as:

$$\Theta(\mathbf{A}) = \frac{1}{2} \sum_{i=1}^{p} \sum_{\substack{e_{j}^{i} \in E_{h}^{i} \\ v_{k}, v_{l} \in \mathbb{V}}} \sum_{\substack{v_{k}, v_{l} \in \mathbb{V} \\ q_{j} \in E_{h}^{n} \\ v_{k}, v_{l} \in \mathbb{V}}} \frac{W_{h_{jj}}^{i} h^{i}(v_{k}, e_{j}^{i}) h^{i}(v_{l}, e_{j}^{i})}{d^{i}(e_{j}^{i})} \|\boldsymbol{\alpha}_{k} - \boldsymbol{\alpha}_{l}\|^{2} \frac{S_{k}}{q_{j}} \frac{S_{k}}{d^{i}(e_{j}^{i})} \|\boldsymbol{\alpha}_{k} - \boldsymbol{\alpha}_{l}\|^{2} \frac{S_{k}}{q_{j}} \frac{S_{k}}{q_{j}}} \frac{S_{k}}{q_{j}} \frac{S_$$

where α_k is the k-th column of the representation matrix **A**. The regularization term $\Theta(\mathbf{A})$ integrates the complex topological structure of data in the subspace representation and promotes similar representations for the pixels within the same hyperedge. We reformulate the regularization in (19) to the following concise form by

$$\Theta(\mathbf{A}) = \sum_{i=1}^{p} \operatorname{tr}(\mathbf{A}\mathbf{L}_{i}\mathbf{A}^{T}) + \operatorname{tr}(\mathbf{A}\mathbf{L}^{n}\mathbf{A}^{T})$$
$$= \operatorname{tr}(\mathbf{A}\mathbf{L}_{h}\mathbf{A}^{T})$$
(20)

where $\mathbf{L}_i = \mathbf{D}_v^i - \mathbf{H}^i \mathbf{W}_h^i (\mathbf{D}_e^i)^{-1} \mathbf{H}^{i^T}$ and $\mathbf{L}^n = \mathbf{D}_v^n - \mathbf{H}^n \mathbf{W}_h^n (\mathbf{D}_e^n)^{-1} \mathbf{H}^{n^T}$ are the Laplacian matrices of the multiscale local hypergraphs \mathcal{G}_l^i and the nonlocal hypergraph \mathcal{G}_n , respectively, and

$$\mathbf{L}_{h} = \sum_{i=1}^{p} \mathbf{L}_{i} + \mathbf{L}^{n}$$
(21)

is the Laplacian matrix corresponding to the hybrid hypergraph \mathcal{G}_h .

C. Optimization Algorithm

According to the definition of $\Theta(\mathbf{A}^t)$ in (20), we derive the following optimization problem from (6):

$$\underset{\mathbf{A}^{t}, \mathbf{Z}, \mathbf{E}^{t}}{\operatorname{arg\,min}} \sum_{t=1}^{T} (\|\mathbf{X}^{t} - \mathbf{X}^{t} \mathbf{A}^{t}\|_{F}^{2} + \lambda_{1} \operatorname{tr}(\mathbf{A}^{t} \mathbf{L}_{h}^{t} \mathbf{A}^{t^{T}}) + \lambda_{2} \|\mathbf{E}^{t}\|_{1}) + \lambda_{3} \|\mathbf{Z}\|_{*}$$
s.t. $\mathbf{A}^{t} = \mathbf{Z} + \mathbf{E}^{t} \quad (\forall t = 1, 2, ..., T)$ (22)

where \mathbf{L}_{h}^{t} is the Laplacian matrix in the *t*-th data source. Directly solving the optimization problem in (22) is difficult and there is no known close-form solution. In this section, we propose an efficient algorithm to solve the optimization problem (22). First, we introduce auxiliaries S^t ($\forall t = 1, 2, ..., T$) and let $\mathbf{A}^t = \mathbf{S}^t$. The problem (22) can be solved equivalently as follows:

$$\underset{\mathbf{A}^{t},\mathbf{S}^{t},\mathbf{Z},\mathbf{E}^{t}}{\operatorname{arg\,min}} \sum_{t=1}^{T} (\|\mathbf{X}^{t} - \mathbf{X}^{t}\mathbf{A}^{t}\|_{F}^{2} + \lambda_{1}\operatorname{tr}(\mathbf{S}^{t}\mathbf{L}_{h}^{t}\mathbf{S}^{t^{T}}) + \lambda_{2}\|\mathbf{E}^{t}\|_{1}) + \lambda_{3}\|\mathbf{Z}\|_{*}$$

s.t. $\mathbf{A}^{t} = \mathbf{Z} + \mathbf{E}^{t}, \ \mathbf{A}^{t} = \mathbf{S}^{t} \ (\forall t = 1, 2, ..., T)$ (23)

We resort to augmented Lagrangian multiplier method and

Algorithm 1 The proposed HMSC model

- 1: Input: $\{\mathbf{X}^t\}_{t=1}^T$, λ_1 , λ_2 , λ_3 , K, p, w, ε
- 2: Laplacian matrices \mathbf{L}_{h}^{t} generation:

for
$$t \leftarrow 1$$
 to T do

- Generate $\mathcal{G}_l^{(t,i)}$ in \mathbf{X}^t (*i*=1,2,...,*p*) by (10) Generate \mathcal{G}_n^t in \mathbf{X}^t by (15)

Compute \mathbf{L}_{h}^{t} by (21)

'7: end for

4: 5:

6:

8: Initialize \mathbf{S}^t , \mathbf{Z} , \mathbf{E}^t , \mathbf{Y}_1^t , \mathbf{Y}_2^t as zero matrices and $\mu = 1$

- while not converged do 9:
- Update \mathbf{A}^t by (26) 10:
- Update S^t by (28) 11:
- 12: Update \mathbf{Z} by (32)
- Update \mathbf{E}^t by (35) 13:
- Update others by (36) 14:
- Check the convergence condition $\|\mathbf{A}^t \mathbf{Z} \mathbf{E}^t\|_{\infty} < \varepsilon$ 15: and $\|\mathbf{A}^t - \mathbf{S}^t\|_{\infty} < \varepsilon$
- 16: end while
- 17: **Output**: Matrix **Z**

solve the following objective function instead:

$$\underset{\mathbf{A}^{t},\mathbf{S}^{t},\mathbf{Z},\mathbf{E}^{t}}{\operatorname{arg\,min}} \sum_{t=1}^{T} (\|\mathbf{X}^{t}-\mathbf{X}^{t}\mathbf{A}^{t}\|_{F}^{2} + \lambda_{1}\operatorname{tr}(\mathbf{S}^{t}\mathbf{L}_{h}^{t}\mathbf{S}^{t^{T}}) \\ + \lambda_{2}\|\mathbf{E}^{t}\|_{1} + \langle \mathbf{Y}_{1}^{t},\mathbf{A}^{t}-\mathbf{Z}-\mathbf{E}^{t}\rangle + \langle \mathbf{Y}_{2}^{t},\mathbf{A}^{t}-\mathbf{S}^{t}\rangle \\ + \frac{\mu}{2}\|\mathbf{A}^{t}-\mathbf{Z}-\mathbf{E}^{t}\|_{F}^{2} + \frac{\mu}{2}\|\mathbf{A}^{t}-\mathbf{S}^{t}\|_{F}^{2}) + \lambda_{3}\|\mathbf{Z}\|_{*}$$

$$(24)$$

where \mathbf{Y}_1^t and \mathbf{Y}_2^t are the Lagrange multipliers introduced for the constraints in (23), respectively and μ is a penalty parameter. Then, we update each of the variables $\{\mathbf{A}^t, \mathbf{S}^t, \mathbf{Z}, \mathbf{E}^t\}$ iteratively by solving one while fixing others based on the alternating direction method of multipliers (ADMM) algorithm.

The subproblem with respect to \mathbf{A}^t (t = 1, 2, ..., T) is formulated by

$$\underset{\mathbf{A}^{t}}{\operatorname{arg\,min}} \|\mathbf{X}^{t} - \mathbf{X}^{t}\mathbf{A}^{t}\|_{F}^{2} + \langle \mathbf{Y}_{1}^{t}, \mathbf{A}^{t} - \mathbf{Z} - \mathbf{E}^{t} \rangle$$
$$+ \langle \mathbf{Y}_{2}^{t}, \mathbf{A}^{t} - \mathbf{S}^{t} \rangle + \frac{\mu}{2} \|\mathbf{A}^{t} - \mathbf{Z} - \mathbf{E}^{t}\|_{F}^{2} + \frac{\mu}{2} \|\mathbf{A}^{t} - \mathbf{S}^{t}\|_{F}^{2}.$$
(25)

We derive the solution by setting the first-order derivative to zero and update \mathbf{A}^t as follows:

$$\mathbf{A}^{t} = (2\mathbf{X}^{t^{T}}\mathbf{X}^{t} + 2\mu\mathbf{I})^{-1}(2\mathbf{X}^{t^{T}}\mathbf{X}^{t} + \mu(\mathbf{Z} + \mathbf{E}^{t} + \mathbf{S}^{t}) - \mathbf{Y}_{1}^{t} - \mathbf{Y}_{2}^{t}).$$
(26)

The objective function with respect to \mathbf{S}^t (t = 1, 2, ..., T) is given by

$$\underset{\mathbf{S}^{t}}{\arg\min} \ \lambda_{1} \operatorname{tr}(\mathbf{S}^{t} \mathbf{L}_{h}^{t} \mathbf{S}^{t^{T}}) + \frac{\mu}{2} \|\mathbf{A}^{t} - \mathbf{S}^{t} + \frac{\mathbf{Y}_{2}^{t}}{\mu}\|_{F}^{2}$$
(27)

The problem in (27) is convex and we update S^t by the following close-form solution:

$$\mathbf{S}^{t} = (\mu \mathbf{A}^{t} + \mathbf{Y}_{2}^{t})(2\lambda_{1}\mathbf{L}_{h}^{t} + \mu \mathbf{I})^{-1}$$
(28)

Next, we solve the subproblem with respect to \mathbf{Z} as shown by

$$\underset{\mathbf{Z}}{\operatorname{arg\,min}} \sum_{t=1}^{T} \frac{\mu}{2} (\|\mathbf{A}^{t} - \mathbf{Z} - \mathbf{E}^{t} + \frac{\mathbf{Y}_{1}^{t}}{\mu}\|_{F}^{2}) + \lambda_{3} \|\mathbf{Z}\|_{*}.$$
(29)

To solve this problem, we first introduce the soft-thresholding operator $\mathcal{D}_{\delta}(\mathbf{X})$. Let $\mathbf{U}\Sigma\mathbf{V}^{T}$ be the singular value decomposition of a matrix \mathbf{X} , i.e., $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^{T}$ where \mathbf{U} and \mathbf{V} are the left and right singular vectors, respectively, and Σ is a diagonal matrix containing the singular values σ_{i} . The operator $\mathcal{D}_{\delta}(\mathbf{X})$ is defined as follows:

$$\mathcal{D}_{\delta}(\mathbf{X}) := \mathbf{U}\mathcal{D}_{\delta}(\mathbf{\Sigma})\mathbf{V}^{T},\tag{30}$$

where $\mathcal{D}_{\delta}(\mathbf{\Sigma}) = \text{diag}(\max((\sigma_i - \delta), 0))$. By reformulating the problem (29) into the following form

$$\underset{\mathbf{Z}}{\operatorname{arg\,min}} \frac{\mu T}{2} \| \frac{1}{T} \sum_{i=1}^{T} (\mathbf{A}^{t} - \mathbf{E}^{t} + \frac{\mathbf{Y}_{1}^{t}}{\mu}) - \mathbf{Z} \|_{F}^{2} + \lambda_{3} \| \mathbf{Z} \|_{*}.$$
(31)

We obtain the following solution of (29):

$$\mathbf{Z} = \mathcal{D}_{\frac{\lambda_3}{\mu T}} \left(\frac{1}{T} \sum_{i=1}^{T} (\mathbf{A}^t - \mathbf{E}^t + \frac{\mathbf{Y}_1^t}{\mu}).$$
(32)

Then, we derive the subproblem with respect to \mathbf{E}^t (t = 1, 2, ..., T) as follows:

$$\underset{\mathbf{E}^{t}}{\operatorname{arg\,min}} \ \lambda_{2} \|\mathbf{E}^{t}\|_{1} + \frac{\mu}{2} \|\mathbf{A}^{t} - \mathbf{Z} - \mathbf{E}^{t} + \frac{\mathbf{Y}_{1}^{t}}{\mu}\|_{F}^{2}.$$
(33)

By introducing the following soft-thresholding operator:

$$\mathcal{R}_{\Delta}(x) = \begin{cases} sgn(x)(|x| - \Delta) & |x| \ge \Delta \\ 0 & otherwise, \end{cases}$$
(34)

we can update \mathbf{E}^t by

$$\mathbf{E}^{t} = \mathcal{R}_{\frac{\lambda_{2}}{\mu}} (\mathbf{A}^{t} - \mathbf{Z} + \frac{\mathbf{Y}_{1}^{t}}{\mu}).$$
(35)

Finally, we update other variables by

$$\begin{aligned} \mathbf{Y}_1^t &= \mathbf{Y}_1^t + \mu(\mathbf{A}^t - \mathbf{Z} - \mathbf{E}^t) \\ \mathbf{Y}_2^t &= \mathbf{Y}_2^t + \mu(\mathbf{A}^t - \mathbf{S}^t) \\ \mu &= \rho\mu, \end{aligned} \tag{36}$$

where ρ is a parameter. Here, we set $\rho = 1.1$ in our algorithm. The complete solver for the problem (22) is summarized in Algorithm 1.

Next, we analyse the computational complexity of updating each variable in Algorithm 1. The time complexity is $\mathcal{O}(TN^3)$ for updating \mathbf{A}^t , $\mathcal{O}(TN^3)$ for updating \mathbf{S}^t and $\mathcal{O}(N^3)$ for updating \mathbf{Z} . The computational complexity to update \mathbf{E}^t , \mathbf{Y}_1^t and \mathbf{Y}_2^t is neglectable. Thus, the overall time complexity of Algorithm 1 is $\mathcal{O}(I(2T+1)N^3)$, where I is the number of iterations.





IV. EXPERIMENTS

We perform experiments on four real data sets to evaluate the effectiveness of the proposed HMSC model. The benchmark methods consist of single-view clustering methods and multi-view clustering methods. The single-view clustering approaches include two classical clustering methods FCM [19] and k-means [18], the clustering by fast search and find of density peaks (CFSFDP) [23], the original SSC model [28], the spatial-spectral ℓ_2 -norm based SSC (L2-SSC) [32] and joint SSC (JSSC) [42]. We are not aware of any reported multi-view clustering methods in remote sensing except the recent work [48], where a spatial-spectral based multi-view low-rank sparse subspace clustering (SSMLC) was proposed. Apart from SSMLC, we also compare our approach to another general multi-view subspace clustering (MSC) method [49], which has achieved the state-of-the-art performance in multiview clustering tasks.

A. Data sets

1) Indian Pines: This hyperspectral image was captured by the Airborne/Visible Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines region in North-western Indiana on June 12, 1992, with 20-m spatial resolution per pixel. The original image size is $145 \times 145 \times 220$. In the experiment, we remove 20 spectral bands in 104-108, 150-163 and 200 due to water absorption. For computational efficiency, a typical subimage with the spatial size 85×70 is tested as in [30, 33, 61], which includes four classes. The false-color composite image and ground truth are shown in Fig. 3 (a) and (b), respectively. To generate a second data source, we utilize an effective spatial feature extended multiattribute profiles (EMAPs) [62], which has demonstrated superior performance in the supervised HSI classification [63]. We follow [62, 63] and consider four different attributes, i.e., area, length of the diagonal, moment of inertia and standard deviation, on the first few principle components of HSI, which are obtained by principle component analysis (PCA). Denote by $\lambda_a, \lambda_d, \lambda_i, \lambda_s$ the parameters which correspond to the four adopted attributes, respectively, we set $\lambda_a = [100, 500, 1000, 5000], \ \lambda_d = [10, 25, 50, 100], \ \lambda_i =$ [0.2, 0.3, 0.4, 0.5] and $\lambda_s = [10, 25, 50, 100]$ as suggested in [62] to extract the multiattribute profiles. We utilize six principle components and concatenate their spatial features, i.e., the multiattribute profiles, as the EMAPs. We refer to [62] for more details.

2) Pavia University: The second HSI was acquired by the Reflected Optics System Imaging Spectrometer (ROSIS) during a flight campaign over Pavia, Northern Italy. The image has a spatial size of 610×340 with 103 spectral bands. A typical area with a size of 200×100 is extracted as the test data which includes eight classes in total. The false color image and ground truth can be found in Fig. 3 (c) and (d). We also employ the EMAPs spatial feature as an additional data source. Specifically, we utilize the same parameter settings as in *Indian Pines* except that the number of principle components is set to four.

3) Salinas_A: This hyperspectral image was collected by the AVIRIS sensor over Salinas Valley, CA, US. The image size is $86 \times 83 \times 204$. The false color image and ground truth are shown in Fig. 3 (e) and (f), which include seven classes of vegetables. We generate the second data source similarly to *Indian Pines* and *Pavia University* by extracting the EMAPs spatial feature of HSI. Two attributes, including area and length of the diagonal, are considered. We generate the EMAPs on the first three principle components of HSI with $\lambda_a = [100, 500, 1000, 5000]$ and $\lambda_d = [10, 25, 50, 100]$.

4) University of Houston (Houston): The last data set contains a HSI and pseudowaveform LiDAR, which is often utilized as a benchmark data set in the task of supervised multimodal classification [64]. The hyperspectral data was acquired by the ITRES-CASI 1500 sensor over the University of Houston campus and the neighbouring urban area. The LiDAR data was collected by an Optech Gemini 280 sensor and then coregistered to the HSI. We select a representative region with the image size of $130 \times 130 \times 144$ as the test data. The false color of HSI and the ground truth of seven classes are shown in Fig. 3 (g) and (h). To increase the discriminative ability of LiDAR, we follow [63] and employ the EMAPs spatial feature of LiDAR as the second data source. The

adopted attributes and parameters are the same as in *Indian Pines*.

B. Experiment setting

We conduct all the experiments in MATLAB on a computer with an Intel[©] core-i7 3930K CPU with 64GB of RAM. The quantitative metrics to measure the clustering performance are overall accuracy (OA), Kappa coefficient (κ) and Normalized Mutual Information (NMI). For a dataset with N samples, the OA is obtained by:

$$OA = \frac{1}{N} \sum_{i=1}^{N} \delta(\operatorname{map}(r_i), l_i), \qquad (37)$$

where r_i is the label of the *i*-th data point obtained by clustering and l_i is the corresponding true label, $\delta(x, y) = 1$ if x = y and is zero otherwise; map(·) is a mapping function that finds the best match between the clustering results and ground truth. We use the Hungarian algorithm to derive the best mapping function. Let $n_{i,j}$ be the number of samples in class *i* that are labeled as class *j*. The Kappa coefficient κ is then defined as:

$$c = \frac{\frac{1}{N} \sum_{i} n_{i,i} - \frac{1}{N^2} \sum_{i} n_{i,+} n_{+,i}}{1 - \frac{1}{N^2} \sum_{i} n_{i,+} n_{+,i}},$$
(38)

where $n_{i,+} = \sum_{j} n_{i,j}$ is the number of samples in class *i* and $n_{+,i} = \sum_{j} n_{j,i}$ is the number of samples that are identified as class *i*. The NMI score is calculated as:

$$\mathbf{NMI} = \frac{\mathbf{I}(l;r)}{\max(\mathbf{H}(l),\mathbf{H}(r))},\tag{39}$$

where I(l; r) denotes the mutual information between l and r, and H(l) and H(r) are their entropies.

For the three evaluation metrics, a larger value indicates a better performance. Note that the label information is only utilized in the evaluation of clustering results. We tune the parameters of all the methods carefully and report their optimal results in terms of OA. For our HMSC method, we fix p = 4and w = 5 in all the data sets.

C. Performance Comparison

Comparative quantitative evaluation of the clustering methods on the four datasets is reported in Tables I-IV and the corresponding clustering maps are shown in Figs. 4-7. We denote by V1 the case where only the original hyperspectral data is utilized, V2 the case where only the second data source is used, and V1+V2 the case where both data sources are applied. In all the tables, the best result is annotated in bold and the second best result is underlined. We report the clustering maps of single-view clustering methods FCM, k-means, CFSFDP, SSC, L2-SSC and JSSC with the data source that produces the best performance. We set the parameters of our method as $\lambda_1 = 5e^{-3}, \lambda_2 = 5e^{-3}, \lambda_3 = 1e^{-2}, K = 15$ for Indian *Pines*, $\lambda_1 = 1e^{-4}$, $\lambda_2 = 5e^{-2}$, $\lambda_3 = 1e^{-5}$, K = 20 for *Pavia* University, $\lambda_1 = 1e^{-1}, \lambda_2 = 5e^{-2}, \lambda_3 = 1e^{-3}, K = 25$ for Salinas_A and $\lambda_1 = 1e^{-3}, \lambda_2 = 5e^{-3}, \lambda_3 = 5e^{-1}, K = 25$ for Houston. The parameters analysis is given in subsection IV-H. Generally, in Tables I-IV the proposed HMSC consistently

9

 TABLE I

 QUANTITATIVE EVALUATION OF DIFFERENT CLUSTERING METHODS FOR THE INDIAN PINES IMAGE:

 CLUSTERING ACCURACY(%), κ coefficient and NMI. (V1: HSI, V2: EMAPS feature)

No.	Class name	F	СМ	k-m	k-means		SFDP	SSC	
		V1	V2	V1	V2	V1	V2	V1	V2
1	Corn-notill	45.87	0.00	43.28	39.90	75.92	32.44	60.00	43.18
2	Grass-trees	99.59	100.00	<u>99.59</u>	86.58	100.00	100.00	98.36	87.81
3	Soybean-notill	59.97	6.97	57.51	2.73	69.40	75.96	76.91	30.60
4	Soybean-mintill	57.95	91.27	63.15	97.9 7	18.61	55.15	50.68	87.68
	OA(%)	62.45	57.78	63.72	66.91	53.72	60.87	65.11	68.00
	κ	0.48	0.38	0.49	0.48	0.40	0.46	0.53	0.53
	NMI	0.42	0.29	0.43	0.51	0.40	0.45	0.41	0.45
No.	Class name	L2-SSC		JS	JSSC		MSC	HMSC	
		V1	V2	V1	V2	V1+V2	V1+V2	V1+V2	
1	Corn-notill	61.09	63.68	74.73	83.78	72.34	61.09	91.24	
2	Grass-trees	99.32	100.00	99.18	100.00	<u>99.59</u>	96.85	100.00	
3	Soybean-notill	79.37	83.20	<u>95.90</u>	95.08	54.51	68.99	100.00	
4	Soybean-mintill	54.89	55.35	86.33	85.34	84.51	62.16	<u>91.53</u>	
	OA(%)	67.78	69.32	87.41	89.05	79.23	68.82	94.28	
	κ	0.5629	0.59	0.82	<u>0.85</u>	0.70	0.57	0.92	
	NMI	0.48	0.53	0.68	0.73	0.53	0.45	0.83	

 TABLE II

 QUANTITATIVE EVALUATION OF DIFFERENT CLUSTERING METHODS FOR THE Pavia University IMAGE:

 CLUSTERING ACCURACY(%), κ COEFFICIENT AND NMI. (V1: HSI, V2: EMAPS FEATURE)

No.	Class name	FC	CM	k-m	eans	CFS	FDP	S	SSC	
		V1	V2	V1	V2	V1	V2	V1	V2	
1	Asphalt	0.00	0.00	0.00	87.53	0.00	0.71	0.00	99.06	
2	Meadows	59.24	97.79	86.07	98.70	62.37	<u>99.87</u>	98.57	<u>99.87</u>	
3	Trees	100.00	9.52	0.00	0.00	0.00	0.00	0.00	0.00	
4	Metal	62.66	56.43	36.88	54.68	35.51	99.85	55.82	98.86	
5	Bare Soil	28.76	41.70	40.17	59.48	58.15	76.44	54.90	67.17	
6	Bitumen	99.77	93.95	<u>99.88</u>	89.88	99.88	100.00	98.84	67.79	
7	Brick	3.19	0.00	36.17	0.00	0.00	100.00	41.49	100.00	
8	Shadows	100.00	<u>99.72</u>	100.00	98.89	100.00	0.00	99.17	99.17	
	OA(%)	51.20	57.94	53.19	69.84	56.70	77.47	64.28	81.33	
	κ	0.43	0.51	0.45	0.64	0.47	0.71	0.57	0.77	
	NMI	0.65	0.71	0.65	0.76	0.67	0.78	0.71	0.86	
No.	Class name	L2-SSC		JS	JSSC		MSC	HMSC		
		V1	V2	V1	V2	V1+V2	V1+V2	V1+V2		
1	Asphalt	0.00	99.76	98.12	99.53	0.00	100.00	100.00		
2	Meadows	94.27	99.87	<u>99.87</u>	100.00	100.00	100.00	100.00		
3	Trees	84.13	90.48	14.29	1.59	0.00	100.00	0.00		
4	Metal	97.34	99.54	96.50	<u>99.70</u>	50.04	99.54	99.54		
5	Bare Soil	51.78	68.19	46.03	81.36	60.73	34.62	75.03		
6	Bitumen	97.21	99.30	96.51	68.26	98.95	98.60	100.00		
7	Brick	39.36	0.00	1.06	100.00	98.94	0.00	100.00		
8	Shadows	<u>99.72</u>	<u>99.72</u>	96.12	0.00	99.45	<u>99.72</u>	100.00		
	OA(%)	71.61	85.59	74.76	81.71	66.45	72.29	89.01		
	κ	0.66	0.82	0.70	0.77	0.60	0.68	0.86		
	NMI	0.74	<u>0.88</u>	0.78	0.84	0.76	0.84	0.90		

achieves the best clustering performance in terms of OA, κ and NMI in the four data sets, demonstrating the effectiveness of our multimodal clustering method. The following observations can be made:

1) The results in Tables I–IV reveal that the subspace clustering methods SSC, L2-SSC, JSSC, SSMLC, MSC and HMSC, outperform the classical clustering methods FCM, k-means and CFSFDP in most cases. Compared with CFSFDP, the most basic subspace clustering method SSC in V1 even achieves OA improvements of 11.39% in *Indian Pines*, 7.58% in *Pavia University*, 4.06% in *Salinas* _A and 16.63% in *Houston*. Compared with FCM and k-means, SSC also yields notable improvements. This demonstrates that subspace clustering is able to uncover better the cluster structure of data than the

 TABLE III

 QUANTITATIVE EVALUATION OF DIFFERENT CLUSTERING METHODS FOR THE Salinas_A IMAGE:

 CLUSTERING ACCURACY(%), κ coefficient and NMI. (V1: HSI, V2: EMAPs feature)

No.	Class name	FC	СМ	k-m	eans	CFSFDP		SSC	
		V1	V2	V1	V2	V1	V2	V1	V2
1	Brocoli-green-weeds-1	99.74	99.74	99.74	0	99.74	99.74	99.23	99.49
2	Corn-senesced-green-weeds	31.12	41.03	28.15	59.79	40.21	39.76	56.59	55.85
3	Lettuce-romaine-4wk	94.97	94.48	94.97	95.62	94.48	100.00	91.40	<u>99.51</u>
4	Lettuce-romaine-5wk	100.00	100.00	100.00	100.00	100.00	100.00	<u>99.93</u>	98.89
5	Lettuce-romaine-6wk	99.26	<u>99.85</u>	99.11	100.00	99.26	100.00	100.00	95.10
6	Lettuce-romaine-7wk	94.62	<u>99.87</u>	94.62	<u>99.87</u>	97.37	100.00	99.12	97.00
	OA(%)	81.21	84.50	80.44	82.07	83.84	84.85	87.90	87.43
	κ	0.77	0.81	0.76	0.78	0.80	0.82	0.85	0.85
	NMI	0.81	0.88	0.80	0.89	0.86	0.89	0.87	0.85
No.	Class name	L2-SSC		JS	JSSC		MSC	HMSC	
		V1	V2	V1	V2	V1+V2	V1+V2	V1+V2	
1	Brocoli-green-weeds-1	99.23	99.49	99.49	98.98	99.49	99.74	99.74	
2	Corn-senesced-green-weeds	58.67	59.94	97.77	<u>99.26</u>	41.55	58.67	100.00	
3	Lettuce-romaine-4wk	95.62	98.38	99.03	98.86	92.69	96.59	100.00	
4	Lettuce-romaine-5wk	99.74	97.70	100.00	100.00	100.00	100.00	100.00	
5	Lettuce-romaine-6wk	<u>99.85</u>	99.41	<u>99.85</u>	100.00	<u>99.85</u>	0.00	100.00	
6	Lettuce-romaine-7wk	99.00	97.12	99.00	99.00	98.37	99.12	100.00	
	OA(%)	88.82	88.56	99.12	<u>99.46</u>	84.18	76.48	99.98	
	κ	0.86	0.86	0.99	<u>0.99</u>	0.81	0.71	1.00	
	NMI	0.88	0.87	0.97	<u>0.98</u>	0.86	0.87	1.00	

 TABLE IV

 QUANTITATIVE EVALUATION OF DIFFERENT CLUSTERING METHODS FOR THE Houston IMAGE:

 CLUSTERING ACCURACY(%), K COEFFICIENT AND NMI. (V1: HSI. V2: LIDAR EMAPS FEATURE)

No.	Class name	FC	CM	k-m	k-means		CFSFDP		SSC	
		V1	V2	V1	V2	V1	V2	V1	V2	
1	Concrete	82.79	46.50	80.70	46.50	78.02	46.50	52.83	46.50	
2	Grass-1	99.65	100.00	100.00	100.00	100.00	100.00	99.30	100.00	
3	Grass-2	0.00	0.00	0.00	0.00	0.00	23.12	71.68	54.48	
4	Parking lot	52.72	67.65	70.89	99.70	43.58	62.01	99.50	62.16	
5	Roof	2.31	95.38	92.31	95.38	0.00	24.62	100.00	0.00	
6	Trees	81.60	51.57	79.42	52.30	1.94	25.18	91.53	57.87	
7	Asphalt	64.15	70.82	95.85	0.00	83.14	92.96	0.00	85.41	
	OA(%)	63.52	61.30	74.97	62.65	56.55	61.16	73.18	64.78	
	κ	0.57	0.53	0.69	0.51	0.48	0.53	0.65	0.58	
	NMI	0.69	0.62	0.73	0.65	0.64	0.62	0.79	0.65	
No.	Class name	L2-SSC		JSSC		SSMLC	MSC	HMSC		
		V1	V2	V1	V2	V1+V2	V1+V2	V1+V2		
1	Concrete	46.50	46.50	53.50	46.50	46.42	53.20	52.19		
2	Grass-1	99.30	100	<u>99.88</u>	100.00	100.00	100	100.00		
3	Grass-2	58.06	41.58	69.89	29.21	97.31	<u>98.57</u>	98.75		
4	Parking lot	99.10	69.10	96.36	96.95	<u>99.50</u>	73.74	90.26		
5	Roof	100.00	0.00	0.00	0.00	93.08	0.00	0.00		
6	Trees	83.05	83.54	84.26	53.51	<u>90.80</u>	81.11	85.96		
7	Asphalt	30.19	65.66	76.35	64.40	0.25	<u>95.09</u>	91.70		
	OA(%)	73.77	65.04	<u>79.54</u>	70.86	74.05	76.92	82.18		
	κ	0.67	0.57	<u>0.75</u>	0.63	0.67	0.72	0.78		
	NMI	0.72	0.61	<u>0.78</u>	0.67	0.80	0.80	0.83		

conventional clustering methods in this task. The superior performance of subspace clustering can be mainly attributed to the efficient subspace representation which learns the correlations of data points in a lower-dimensional subspaces by solving self-representation based optimization problems. 2) The methods that use spatial regularization show indeed improvement over the methods based on spectral information alone. For instance, in the single-view subspace clustering methods, L2-SSC and JSSC obtain the OA gains of (2.67%, 22.3%) with hyperspectral data (V1), and (1.32%, 21.05%)



Fig. 4. *Indian Pines* image. (a) Ground truth and clustering maps of (b) FCM, (c) k-means, (d) CFSFDP, (e) SSC, (f) L2-SSC, (g) JSSC, (h) SSMLC, (i) MSC, (j) HMSC.



Fig. 5. *Pavia University* image. (a) Ground truth and clustering maps of (b) FCM, (c) k-means, (d) CFSFDP, (e) SSC, (f) L2-SSC, (g) JSSC, (h) SSMLC, (i) MSC, (j) HMSC.

with generated feature (V2) in the Indian Pines. The OA enhancements can be also observed in other three data sets. In the multi-view subspace clustering methods, HMSC outperforms SSMLC and MSC and produces higher OA, κ and NMI. Specifically, Compared with SSMLC, HMSC yields the improvements of OA by 15.05%, 22.56%, 15.8% and 8.13%, respectively, in Indian Pines, Pavia University, Salinas _A and Houston. Compared with MSC, we obtain better cluster performance with OA gains of 25.46%, 16.72%, 23.5% and 5.26%, respectively, in Indian Pines, Pavia University, Salinas A and Houston. These observations verify the importance of incorporating the spatial information of data in subspace clustering methods, which promotes the spatial correlations in the subspace representations. Although the multi-view clustering methods SSMLC and MSC cluster data by using multiview data, the spatial correlations of pixels in each view are not taken into consideration, resulting in the unsatisfactory performance. We utilize both the local and nonlocal spatial information and integrate them elegantly into a hybridhypergraph based regularization, achieving significant perfor-



Fig. 6. *Salinas_A* image. (a) Ground truth and clustering maps of (b) FCM, (c) k-means, (d) CFSFDP, (e) SSC, (f) L2-SSC, (g) JSSC, (h) SSMLC, (i) MSC, (j) HMSC.



Fig. 7. *Houston* image. (a) Ground truth and clustering maps of (b) FCM, (c) k-means, (d) CFSFDP, (e) SSC, (f) L2-SSC, (g) JSSC, (h) SSMLC, (i) MSC, (j) HMSC.

mance improvements. *Indian Pines* is known as a challenging hyperspectral data in classification task due to noise, large intra-class spectral variability and high inter-class similarity [65], which results in the poor performance (often OA is less than 70%) in most of the clustering algorithms. However, our multimodal clustering method produces far better performance than the competitors with overall accuracy of 94.28% and class-specific accuracies of 91.24%, 100%, 100% and 91.53% in Table I.

3) Our multi-view clustering method HMSC shows better clustering performance in terms of OA, κ and NMI than that of all the single-view clustering methods in the four datasets. The single-view clustering methods can obtain significantly different results in the two data sources. For instance, the OA of SSC in V1 is 64.28% in Pavia University while the OA of SSC in V2 is 81.33%. This is caused by the diverse information in V1 and V2. The data V1 (HSIs) provides rich spectral information while the data V2 captures the important spatial information of HSIs in Indian Pines, Pavia University and Salinas A or altitude information of the ground objects in Houston. In the Indian Pines and Pavia University datasets, most of the singleview clustering methods, including FCM, k-means, CFSFDP, SSC, L2-SSC and JSSC, yield better performance in V2 than that in V1, which demonstrates the superior discriminability of spatial feature. In the Salinas_A data set, all the singleview clustering methods perform similarly in both data sources with comparable clustering results. In the Houston dataset, the spectral information captured in V1 is able to differentiate the ground objects better than the altitude information in



Fig. 8. The effect of the local and nonlocal hypergraph regularization in the four data sets.



Fig. 9. The effect of multi-scale local hypergraphs in the four data sets.

V2, which is evidenced by the superior performance in V1. The proposed HMSC exploits the complementary information from both data sources, and thus produces excellent results that outperform all the single-view clustering methods in both data sources. The multi-view clustering methods SSMLC and MSC also employ both data sources. However, their clustering performance is less stable and sometimes shows lower accuracy than the classical methods such as the results in *Pavia University* and *Salinas_A*.

4) The clustering maps in Figs. 4–7 present the visual results of all the clustering methods in the four data sets, respectively. Note that for each method the clustering map using the data source that produces the highest accuracy is reported. The results reveal that the clustering maps of HMSC always match the best to the corresponding ground truths, which is consistent to the quantitative evaluations in Tables I–IV. The clustering methods FCM, k-means, CFSFDP, SSMLC and MSC treat each data points independently without considering the important spatial correlations between data points during the clustering process, and thus cannot preserve well the homogeneity of ground objects, resulting in the deteriorated clustering maps that are significanlty affected by impulse noise. Such effect is less pronounced in the clustering maps of L2-SSC, JSSC and



Fig. 10. Performance comparisons between using normal graph and hypergraph in the proposed model.



Fig. 11. The performance of using hybrid-hypergraph in single data source.

HMSC due to the exploitation of spatial constraints, which effectively increases the spatial dependencies among similar data points in the clustering model.

D. The effect of local and nonlocal hypergraph regularization

The superiority of our model partially benefits from the introduction of the proposed hybrid-hypergraph regularization, where the hybrid hypergraph consists of a series of local spatial hypergraphs and spatial-wise nonlocal hypergraph. We analyse here the effect of local and nonlocal hypergraph regularization to the overall accuracy of HMSC. In this subsection, we conduct experiments by adopting only one type of hypergraph, i.e., local or nonlocal hypergraphs in the HMSC, and report the overall accuracies in Fig. 8. The experimental results reveal that the local spatial regularization results in much higher accuracies than the nonlocal spatial regularization in Indian Pines, Pavia University and Salinas A. The derived OAs also outperform most of the clustering methods in Tables I-II. While in the Houston, the nonlocal spatial regularization plays the dominant roles and yield better clustering results than that with local hypergraph regularization. The results in Fig. 8 conclude that the hybrid-hypergraph regularization yields consistently the best performance in the three experimental



Fig. 12. The effect of λ_1 , λ_2 and λ_3 in HMSC to the overall accuracy in the four data sets (the x-axis is in log_{10}).



Fig. 13. The effect of K in the nonlocal hypergraph to the overall accuracy.

settings (local, nonlocal and hybrid), which demonstrates the effectiveness of the hybrid hypergraph in our model.

E. The effect of multi-scale local hypergraphs

The adaptive super-pixel segmentations provide four segmentation maps at four scales, which are further utilized to construct the multi-scale local spatial hypergraphs. The aforementioned analysis indicates the superiority of using the multiscale local hypergraphs. We investigate here the contribution of local hypergraph regularization at each scale to the overall accuracy of our model in the four data sets by fixing nonlocal hypergraph. We denote by Scale-1, Scale-2, Scale-3 and Scale-4 the cases using local hypergraphs in the HMSC model with segmentation numbers n_1, n_2, n_3 and n_4 , respectively, and Multi-scale the case using all the local hypergraphs. The results in Fig. 9 suggest that the Scale-1 hypergraph always results in superior performance in terms of OA over the other three scales, and a higher level segmentation (more fine segmentation) leads to a reduced clustering performance. This is reasonable because finer segmentation results in more superpixels and each super-pixel may only include parts of the pixels belonging to a ground object, which dilutes the local spatial correlations to a certain degree. The results in Fig. 9 also show that the combined multi-scale local hypergraph regularization yields the best result, which is mainly attributed to the joint contribution of the multi-scale local hypergraphs, demonstrating the efficacy of our approach.

F. The effect of hypergraph

To study the influence of hypergraph on the performance of our method, we conduct experiments by replacing the adopted hypergraph with a normal graph where the weights of connected edges between nodes *i* and *j* are calculated by $W_{i,j} = \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma})$. The connectivities of edges in normal graph are obtained by using the same super-pixel segmentation and KNN results as in the hypergraph. The clustering results are shown in Fig. 10. Generally, hypergraph results in higher accuracies than the normal graph in our model. The OA improvements using hypergraph are significant in the data sets *Indian Pines*, *Pavia University* and *Houston*, and in *Salinas_A*, results are comparable, demonstrating the benefit of using hypergraph. It should be noted that even with normal graph our model outperforms the compared methods in most cases as seen in Tables I–IV.

G. The performance in single data source

We conduct experiments in this subsection to evaluate the performance of using hybrid-hypergraph regularization with single data source. In the single-source setting, we aim to minimize the objective function $\|\mathbf{X}^t - \mathbf{X}^t \mathbf{A}^t\|_F^2 + \lambda_1 \operatorname{tr}(\mathbf{A}^t \mathbf{L}_h^t {\mathbf{A}^t}^T)$. The clustering results for four data sets are shown in Fig. 11, where V1 and V2 denote the clustering result with HSI and generated features, respectively, and V1+V2 is the multimodal clustering result. In Fig. 11, we obtain the same conclusions as discussed in the previous part. Firstly, the accuracy with V2 is much higher than that with V1 in Indian Pines and Pavia University, showing the spatial feature is more discriminative than the original spectral information. Secondly, in Salinas_A both data sources perform similarly. Thirdly, the hyperspectral data differentiates the ground objects better than the LiDAR data in Houston. Lastly, the multimodal clustering result outperforms the results using single data source, which verifies the benefit of our multimodal clustering model. Moreover, we also observe that the accuracies obtained in each data source often exceed the accuracies of the competitors in Tables I-IV, which mainly benefits from the adopted hybrid-hypergraph regularization.

H. Parameter analysis

To analyse the effect of the parameters $\lambda_1, \lambda_2, \lambda_3$ and K to the clustering performance, we tune one parameter while fixing



Fig. 14. The convergence curves of the proposed model. From left to right: Indian Pines, Pavia University, Salinas_A and Houston

 TABLE V

 THE RUNNING TIME OF DIFFERENT CLUSTERING METHODS (IN SECONDS)

Data sets	FCM	k-means	CFSFDP	SSC	L2-SSC	JSSC	SSMLC	MSC	HMSC
Indian Pines	3.5	0.4	4.5	234	256	107	2999	15250	1149
Pavia University	5.8	0.4	7.0	677	6649	323	8888	46657	3461
Salinas_A	0.6	0.1	6.2	397	592	202	5221	7216	2291
Houston	5.8	0.5	<u>3.7</u>	932	11859	237	6852	24434	1192

 TABLE VI

 THE CLUSTERING PERFORMANCE OF HMSC FOR THE Indian Pine IMAGE

No.	Data sources	OA	κ	NMI	Time (in seconds)
1	HSI+Gabor	88.41	0.84	0.72	1362
2	HSI+EMAPs	94.28	0.92	0.83	1149
3	HSI+RSF	93.26	0.90	0.80	1473
4	HSI+EMAPs+Gabor	94.63	0.92	0.84	1460
5	HSI+EMAPs+Gabor+RSF	95.35	0.93	0.86	1733
3	HSI+EMAPS+Gabor+RSF	95.55	0.95	0.80	1/55

others in each experiment, and report the resulting overall accuracy in the four data sets. The results with respect to λ_1, λ_2 and λ_3 are shown in Fig. 12 and the results with varying K are shown in Fig. 13. In general, it is observed that the performance are quite stable to all the parameters in a certain range, which indicates that our algorithm is insensitive to the parameter perturbation. Moreover, the results show significant performance changes in a global view, indicating the efficacy of the introduced regularizations. Notably, compared with the base line of SSC (OA=68%, 81.33%, 87.96% and 73.18% in *Indian Pines, Pavia University, Salinas_A* and *Houston*, respectively), most of the results from Figs. 12 and 13 yield improved performance, which verifies the superiority of our multimodal clustering model.

I. Convergence analysis

To investigate the convergence performance, we conduct experiments for our method in the four data sets and plot the values of objective function with respect to the iteration numbers in Fig. 14. The results show that our method mostly converges in thirty iterations, confirming the practical convergence of our optimization algorithm.

J. Running time

We show in Table V the running time of different clustering methods on the four data sets. The reported running time corresponds to the whole clustering procedure for each method. For the single-view clustering methods, the running time corresponding to the better result among two views is reported. The results show that k-means, FCM and CFSFDP are more efficient than the tested subspace clustering methods. Compared to SSC, L2-SSC and JSSC, multi-view clustering methods SSMLC, MSC and HMSC often use more running time. Among the multi-view clustering methods, our HMSC is the fastest method.

K. Clustering with different data sources

We show the clustering results of HMSC on the data set Indian Pines by using different data sources in Table VI. In addition to EMAPs, two more features of HSI are generated as new views of the same scene, including Gabor feature and a representation-based spatial feature (RSF) extracted by [66]. We utilize two-dimensional Gabor filters [67] with two scales and four orientations on each of the first few PCs of HSI, which contains 99% of the total variation of the data. The filtered PCs are concatenated as the Gabor feature. RSF is obtained by solving a representation-based optimization problem. We set the feature dimension of RSF to 200. Other parameters are set according to [66]. The results in Table VI show that more data sources yield better results in terms of OA, κ and NMI. The price is slight increases of running time, which is acceptable. Among the results using two data sources, the combination of HSI and EMAPs produces the highest accuracy. Moreover, we observe that the results in Table VI outperform all the single-view clustering results in Table I that utilize HSI alone.

V. CONCLUSION

In this paper, a novel multi-view subspace clustering method is proposed for the clustering of hyperspectral images. We design a new data fusion model to leverage the complementary information from multi-view data. This model decomposes the subspace representation in each view as a combination of a global consensus low-rank matrix (shared by all the views) and a view-specific sparse matrix, which enables the learning of a desirable cluster structure in the low-dimensional subspaces with flexible tolerance to the data geometric structure in each view. We also propose a manifold-based spatial regularization by embedding a hybrid hypergraph in the subspace representation. The hybrid hypergraph consists of a series of multiscale local hypergraphs, defined by the super-pixel segmentations at different scales, and a nonlocal hypergraph which is constructed by the K nearest neighbours of each centralized square patch. Thus, the proposed regularization encodes effectively the local and nonlocal high-order spatial context of data. We developed an efficient algorithm to solve the resulting model. Extensive experiments on four real data sets demonstrate favourable performance of our method compared to the current state-of-the-art in the field.

VI. ACKNOWLEDGEMENT

The authors would like to thank the Hyperspectral Image Analysis Group and the NSF-Funded Center for Airborne Laser Mapping, University of Houston, for providing the *University of Houston* image.

REFERENCES

- [1] M. A. Lee, Y. Huang, H. Yao, S. J. Thomson, and L. M. Bruce, "Determining the effects of storage on cotton and soybean leaf samples for hyperspectral analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2562–2570, 2014.
- [2] H. Zhang, J. Kang, X. Xu, and L. Zhang, "Accessing the temporal and spectral features in crop type mapping using multitemporal sentinel-2 imagery: A case study of Yi'an county, Heilongjiang province, China," *Comput. Electron. Agric.*, vol. 176, p. 105618, 2020.
- [3] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, 2009.
- [4] N. Fox, A. Parbhakar-Fox, J. Moltzen, S. Feig, K. Goemann, and J. Huntington, "Applications of hyperspectral mineralogy for geoenvironmental characterisation," *Minerals Engineering*, vol. 107, pp. 63–77, 2017.
- [5] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, 2014.
- [6] H. Zhang, Y. Song, C. Han, and L. Zhang, "Remote sensing image spatiotemporal fusion using a generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, 2020.
- [7] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectralspatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, 2017.
- [8] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, 2017.
- [9] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, 2019.
- [10] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, 2005.

- [11] S. Jia, J. Zhuang, L. Deng, J. Zhu, M. Xu, J. Zhou, and X. Jia, "3-D Gaussian–Gabor feature extraction and selection for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8813–8826, 2019.
- [12] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973– 3985, 2011.
- [13] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learningbased classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, 2014.
- [14] P. Ghamisi, E. Maggiori, S. Li, R. Souza, Y. Tarablaka, G. Moser, A. De Giorgi, L. Fang, Y. Chen, M. Chi *et al.*, "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, 2018.
- [15] A. Ghosh, N. S. Mishra, and S. Ghosh, "Fuzzy clustering algorithms for unsupervised change detection in remote sensing images," *Information Sciences*, vol. 181, no. 4, pp. 699–715, 2011.
- [16] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.
- [17] J. W. Lee, D. S. Yeung, and E. C. Tsang, "Hierarchical clustering based on ordinal consistency," *Pattern Recognit.*, vol. 38, no. 11, pp. 1913–1925, 2005.
- [18] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982.
- [19] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," 1981.
- [20] 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. Syst.*, vol. 34, no. 4, pp. 1907–1916, 2004.
- [21] 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, 2013.
- [22] S. Niazmardi, S. Homayouni, and A. Safari, "An improved FCM algorithm based on the SVDD for unsupervised hyperspectral data classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 831–839, 2013.
- [23] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [24] 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, 2011.
- [25] R. Vidal, "Subspace clustering," IEEE Signal Process. Mag., vol. 28, no. 2, pp. 52–68, 2011.
- [26] T. Zhang, A. Szlam, Y. Wang, and G. Lerman, "Hybrid linear modeling via local best-fit flats," *Int. J. Comput. Vision*, vol. 100, no. 3, pp. 217–240, 2012.
- [27] 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.
- [28] 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.
- [29] J. Wen, B. Zhang, Y. Xu, J. Yang, and N. Han, "Adaptive weighted nonnegative low-rank representation," *Pattern Recognition*, vol. 81, pp. 326–340, 2018.
- [30] 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.

- [31] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Hyperspectral image clustering: Current achievements and future lines," *IEEE Geosci. Remote Sens. Mag.*, 2020.
- [32] 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.
- [33] S. Huang, H. Zhang, and A. Pižurica, "Joint sparsity based sparse subspace clustering for hyperspectral images," in *Proc. IEEE ICIP*, 2018, pp. 3878–3882.
- [34] 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.*, no. 99, pp. 1–15, 2018.
- [35] S. Huang, H. Zhang, Q. Du, and A. Pizurica, "Sketch-based subspace clustering of hyperspectral images," *Remote Sens.*, vol. 12, no. 5, 2020.
- [36] 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, p. 335, 2017.
- [37] Q. Yan, Y. Ding, J.-J. Zhang, Y. Xia, and C.-H. Zheng, "A discriminated similarity matrix construction based on sparse subspace clustering algorithm for hyperspectral imagery," *Cognitive Systems Research*, vol. 53, pp. 98–110, 2019.
- [38] R. Wang, F. Nie, and W. Yu, "Fast spectral clustering with anchor graph for large hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2003–2007, 2017.
- [39] R. Wang, F. Nie, Z. Wang, F. He, and X. Li, "Scalable graph-based clustering with nonnegative relaxation for large hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, 2019.
- [40] S. Huang, H. Zhang, and A. Pižurica, "Landmark-based largescale sparse subspace clustering method for hyperspectral images," in *Proc. IEEE IGARSS*, 2019, pp. 799–802.
- [41] Q. Yan, Y. Ding, Y. Xia, Y. Chong, and C. Zheng, "Class probability propagation of supervised information based on sparse subspace clustering for hyperspectral images," *Remote Sensing*, vol. 9, no. 10, p. 1017, 2017.
- [42] 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.
- [43] Y. Wan, Y. Zhong, A. Ma, and L. Zhang, "Multi-objective sparse subspace clustering for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, 2019.
- [44] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal lowrank tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3071–3084, 2020.
- [45] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, 2005.
- [46] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, 2015.
- [47] S. Jia, L. Shen, J. Zhu, and Q. Li, "A 3-D Gabor phase-based coding and matching framework for hyperspectral imagery classification," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1176– 1188, 2017.
- [48] 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.
- [49] Y. Xie, W. Zhang, Y. Qu, L. Dai, and D. Tao, "Hyper-Laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 572–586, 2018.

- [50] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, 2018.
- [51] L. Huang, H.-Y. Chao, and C.-D. Wang, "Multi-view intact space clustering," *Pattern Recognit.*, vol. 86, pp. 344–353, 2019.
- [52] X. Li, H. Zhang, R. Wang, and F. Nie, "Multi-view clustering: A scalable and parameter-free bipartite graph fusion method," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020.
- [53] Q. Liu, Y. Sun, C. Wang, T. Liu, and D. Tao, "Elastic net hypergraph learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 452–463, 2016.
- [54] 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.
- [55] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5923–5936, 2014.
- [56] B. Song, J. Li, M. Dalla Mura, P. Li, A. Plaza, J. M. Bioucas-Dias, J. A. Benediktsson, and J. Chanussot, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, 2013.
- [57] J. Wen, X. Fang, Y. Xu, C. Tian, and L. Fei, "Low-rank representation with adaptive graph regularization," *Neural Networks*, vol. 108, pp. 83–96, 2018.
- [58] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [59] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Nonlocal means regularized sketched reweighted sparse and low-rank subspace clustering for large hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, 2020.
- [60] J. Xue, Y. Q. Zhao, Y. Bu, W. Liao, J. C. W. Chan, and W. Philips, "Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3084–3097, 2021.
- [61] Y. Pan, Y. Jiao, T. Li, and Y. Gu, "An efficient algorithm for hyperspectral image clustering," in *Proc. IEEE ICASSP*, 2019, pp. 2167–2171.
- [62] 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.
- [63] 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.
- [64] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pižurica, S. Gautama et al., "Hyperspectral and LiDAR data fusion: Outcome of the 2013 grss data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, 2014.
- [65] S. Zhang, J. Li, K. Liu, C. Deng, L. Liu, and A. Plaza, "Hyperspectral unmixing based on local collaborative sparse regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 631–635, 2016.
- [66] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6976– 6985, 2016.
- [67] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *JOSA A*, vol. 2, no. 7, pp. 1160–1169, 1985.



Shaoguang Huang (S'17-M'19) received the M.S. degree in telecommunication and information system from Shandong University, Jinan, China, in 2015 and the Ph.D. degree in computer science engineering from Ghent University, Belgium, in 2019.

He is currently a Post-Doctoral Research Fellow with the Group for Artificial Intelligence and Sparse Modelling (GAIM), Ghent University, Belgium. His area of interests includes image processing, sparse representation, clustering, hyperspectral image analvsis and machine learning.

He serves as a Guest Editor for the Remote Sensing. He is a Reviewer for the international journals IEEE Transactions on Geoscience and Remote Sensing, IEEE Transactions on Circuits and Systems for Video Technology, IEEE Transactions on Biomedical Circuits and Systems, Remote Sensing, Computers and Geosciences, etc.



Hongyan Zhang (M'13-SM'16) received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 2005 and 2010, respectively.

Dr. Zhang has been currently a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, since 2016. He is a Young Chang-Jiang Scholar appointed by the Ministry of Education of China. He scored 1st in the IEEE GRSS Data Fusion

Contest of 2019 and 2021 organized by the IEEE Image Analysis and Data Fusion Technical Committee. He has authored/co-authored more than 100 research papers and 8 patents. His research interests include image reconstruction for quality improvement, hyperspectral information processing and agricultural remote sensing.

Dr. Zhang serves as an Associate Editor of Photogrammetric Engineering & Remote Sensing and Computers & Geosciences and the Session Chair of the 2016 IEEE IGARSS conference and the 2015 IEEE WHISPERS conference. He is a Reviewer of more than 30 international academic journals, including IEEE Transactions on Geoscience and Remote Sensing, IEEE Transactions on Image Processing, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Geoscience and Remote Sensing Letters and so on.



Aleksandra Pižurica (M'03-SM'15) received the Diploma in electrical engineering from the University of Novi Sad, Serbia, in 1994, the Master of Science degree in telecommunications from the University of Belgrade, Serbia, in 1997, and the Ph.D. degree in engineering from Ghent University, Belgium, in 2002. She is a Professor in statistical image modeling with Ghent University. Her research interests include the area of signal and image processing and machine learning, including multiresolution statistical image models, Markov Random

Field models, sparse coding, representation learning, and image and video reconstruction, restoration, and analysis.

Prof. Pižurica served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2012 – 2016), Senior Area Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2016 – 2019) and an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016 – 2020). She was also the Lead Guest Editor for the EURASIP Journal on Advances in Signal Processing for the Special Issue "Advanced Statistical Tools for Enhanced Quality Digital Imaging with Realistic Capture Models" (2013). The work of her team has been awarded twice the Best Paper Award of the IEEE Geoscience and Remote Sensing Society Data Fusion contest, in 2013 and 2014. She received the scientific prize "de Boelpaepe" for 2013–2014, awarded by the Royal Academy of Science, Letters and Fine Arts of Belgium for her contributions to statistical image modeling and applications to digital painting analysis.