

# A Survey on Superpixel Segmentation as a Pre-Processing Step in Hyperspectral Image Analysis

Subhashree Subudhi, *Student Member, IEEE*, Ram Narayan Patro, *Student Member, IEEE*, Pradyut Kumar Biswal\*, *Senior Member, IEEE*, and Fabio Dell'Acqua, *Senior Member, IEEE*

**Abstract**—Recent developments in hyperspectral sensors have made it possible to acquire HyperSpectral Images (HSI) with higher spectral and spatial resolution. Hence, it is now possible to extract detailed information about relatively smaller structures. Despite these advantages, HSI suffers from many challenges also, like higher spatial variability of spectral signatures, the Hughes effect due to higher dimensionality, and a limited number of labeled training samples compared to the dimensions of the spectral space. Superpixels can be a potentially effective tool in tackling these challenges. Superpixel segmentation is a process of segmenting the spatial image into several semantic sub-regions with similar characteristic features. Such grouping by similarity can significantly ease the subsequent processing steps. Because of this, superpixels have been successfully applied to various fields of HSI processing like Classification, Spectral Unmixing, Dimensionality Reduction, Band Selection, Active Learning, Denoising, and Anomaly Detection. This paper focuses on Classification, presenting a detailed survey of superpixel segmentation approaches for the classification of HSI. The Superpixel creation algorithm framework and post-processing frameworks for superpixels in HSI are also analyzed. Also, a brief description of various application areas of superpixels is provided. An experimental analysis of existing superpixel segmentation approaches is also provided in this paper, supported by quantitative results on standard benchmark datasets. The challenges and future research directions for the implementation of superpixel algorithms are also discussed.

**Index Terms**—Hyperspectral Image, Superpixel Segmentation, Evaluation.

## I. INTRODUCTION

In recent years, HyperSpectral (HS) image analysis has gained significant attention in the remote sensing context, because of improvements in quality and availability of data. Indeed, HS data may carry a large volume of information, as each pixel densely samples the spectral response of the object contained in it. Such response, called the “signature” of the object, is unique to the material or mix of materials composing the object. The practical importance of the spectral signature concept is testified by the appearance of special HS sensors with the capacity to recognize distinguishing features that are then used for signature-based compression [1], [2].

Ram Narayan Patro, Subhashree Subudhi and Pradyut Kumar Biswal\* are with the Department of Electronics and Communication Engineering, IIIT, Bhubaneswar, India

Fabio Dell'Acqua is with the Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

Manuscript received Jan 20, 2018; revised xxxxx

When HSI data collection is intended for classification and mapping purposes, as it is frequently the case, Ground Truth (GT) information is needed to train the classifiers and perform accuracy assessments.

Generating reliable GT can be a challenging and expensive endeavour: visual interpretation can suffer from ambiguities, whereas direct inspection may be costly and possibly difficult due to accessibility issues. In any case, GT generation is a time- and resource-consuming process. Hence, it is often suggested that solutions be sought in classification techniques that are capable of:

- utilizing unlabelled samples, and
- predicting the labels of neighboring locations

When the unlabelled neighboring pixel information is also taken into consideration, the burden on the training sample collection can be significantly reduced [3]. Recently, indeed, this type of approach leveraging spatial information - or the various forms of correlation between spatially adjacent samples - has been increasingly adopted in several aspects of HyperSpectral Image (HSI) processing like Anomaly Detection, Band Selection, Unmixing, and Classification.

Considering specifically the field of classification, traditional classifiers (e.g.  $K$ -nearest neighbors or KNN [4], SVM [4], as well as neural-network-based ones [5], etc.), when they incorporate spectral features only may often be biased by atmospheric disturbances and noise. This creates a labeling uncertainty and salt-and-pepper noise in the classification map [6]. Often, this is addressed either with inclusion of spatial features into classification [7], or post classification voting strategies [8]. In general, the voting strategies rely on window based neighborhood operation. In this latter case, however, the selection of window size is tricky and cannot be operated with a universal assumption [9].

Superpixels have been found to offer a possible solution to this and other problems in HSI data analysis, thanks to their capability to change shape and size in relation with spatial information contained in the hyperspectral scene. Indeed, superpixels may also help when incorporated in post-classification voting strategies [10].

Superpixels, due to their inherent properties have gained a lot of attention since they were first named and introduced in 2003 [11]. The concept was quickly adopted in a wide range of applications. Several superpixel segmentation algorithms available in literature are specifically designed for natural

images. In [12], a comprehensive evaluation of 28 state-of-the-art superpixel algorithms are provided. Also, in [13], the performance of SLIC algorithm is compared with other state-of-the-art methods. These survey works mainly focuses on the superpixel segmentation algorithms. Recently, superpixels have gained a lot of popularity in the field of HSI processing. It is desirable to incorporate superpixel segmentation for HSI processing due to the powerful capability of superpixels to adapt to the spatial structure of the depicted objects and group pixels into spatially meaningful clusters. Researchers have successfully applied superpixels in various applications like classification [9], spectral unmixing [14], dimensionality reduction [15], band selection [16], active learning [17], denoising [18], and anomaly detection [19] in hyperspectral images. Frequently, the same algorithms that were developed for natural images are also utilized for performing superpixel segmentation. There exist no dedicated superpixel segmentation algorithms that were specifically designed for HSI, as probably those mutated from natural image processing perform satisfactorily in the new domain. In this survey, for the first time the authors have focused on the different ways in which known superpixel approaches can be profitably used as a pre-processing step for HSI analysis, with special regards to classification.

To demonstrate the increasing popularity of superpixel, in Figure. 1 a plot showing the number of available articles in IEEE Xplore on the subject "Superpixel segmentation for HSI" during the period 2010 to 2020 is provided. It can be observed that the number of papers dedicated to this subject has increased drastically in recent years. Figure 2 contains a pie chart indicating the percentage of articles using superpixel segmentation in various application areas of HSI processing published in IEEE during the same period; the reader's attention is drawn to the relative weight of classification, but also to the range of possible other applications addressed. Due to such widespread applicability of superpixels in the field of HSI processing in this work, a detailed survey of various superpixel segmentation approaches is provided for the reader's reference.

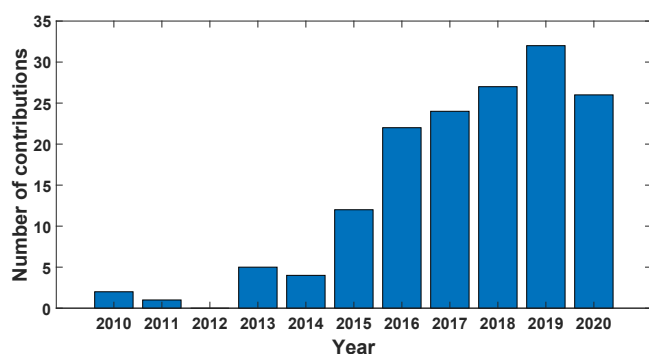


Fig. 1. The number of articles available in IEEE Xplore on the subject Superpixel segmentation for HSI during the period 2010 to 2020

The organization of the rest of the manuscript is as follows. In section II, the concept of superpixel (especially as opposed to that of a pixel) is introduced and discussed. Then, in Section III, the effect of dimension on superpixels is described. A

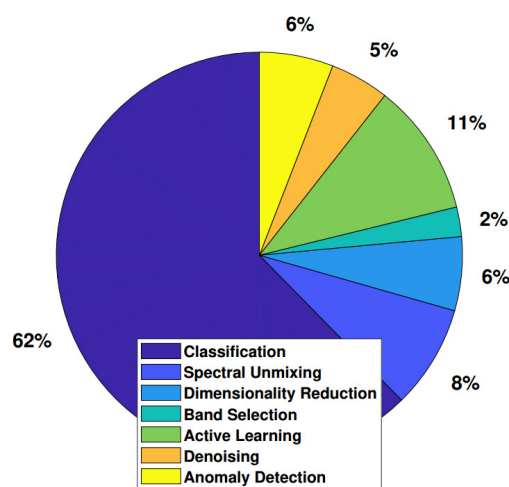


Fig. 2. Application of superpixels in different applications for HSI

superpixel creation algorithm framework in HSI is explained in Section IV. Next, various post-processing frameworks for superpixels are provided in section V. In Section VI, various applications of superpixels are explained. Open challenges are discussed in Section VIII. Finally, some conclusions and future research lines are presented in Section IX. The basic notations used in the paper are provided in Section A in Appendix.

## II. FROM PIXELS TO SUPERPIXELS

In the process of HSI image formation, each elementary piece of information or pixel reports a measured amount of incoming energy, referring to the observed object. Although from a technological standpoint this makes perfect sense, in the context of Earth observation the hyperspectral data generated in this way suffer from two inherent limitations: 1) Pixels are basically the result of spatial discretization 2) in extensive images, a large number of samples, in both spatial and spectral dimensions, makes many information extraction algorithms computationally unfeasible [11]. To overcome the aforementioned problems, Ren and Malik introduced the concept of superpixel segmentation [11].

Superpixels, which may serve as a preparation step to image segmentation, can be defined as an unsupervised over-segmentation of an image into several semantic sub-regions, bearing similar characteristic features. Using superpixels for segmentation has several advantages, like:

- Features can be computed on more meaningful regions instead of acting on the basis of individual pixels.
- thanks to superpixel segmentation, the computational complexity downstream reduces drastically as the input entries for subsequent algorithms are down-scaled from the order of magnitude of pixels to that of regions, without losing significant information.

The benefits of using superpixels over pixels include:

- **Spatial information:** Superpixels provide a robust way to exploit spatial contextual information as they group spectrally similar regions [20].

- **Noise robustness:** they can extract potential low-dimensional features even under noisy conditions [15].
- **Pseudo-label generation:** As GT is not available for all the pixels in HSI, superpixels can be utilized to generate pseudo-labels. Pixels within the same superpixel are spatially and spectrally related so their labels may be propagated [21].
- **Impact on Active Learning:** In active learning, the most informative unlabeled samples can be selected by using the spatial contextual information provided by superpixels [17].
- **Lower computational complexity:** processing steps are applied to regions instead of each individual pixel; this opens the way to more efficient processing [22].
- **Preservation of Boundary information:** Superpixels can very well preserve the boundary or edge features in an image [23], and consequently, the related spatial information is conserved.

In order to be useful, the generated superpixels must possess certain properties, listed below:

- **Homogeneity:** The generated superpixels must have "uniform" pixel values.
- **Boundary Adherence:** Superpixel boundaries must match the object boundaries.
- **Regularity:** Superpixels must form a regular pattern in the image
- **Time Complexity:** The generated Superpixels should have lower computational complexity.
- **Connected Partition:** Superpixels consist of a connected set of pixels and do not overlap with each other.

### III. DIMENSION AND SUPERPIXEL

A basic definition for the objective of superpixel segmentation algorithms can be given as that of grouping together clusters of spatially adjacent pixels with similar spectral features. As a consequence, all pixels within a given superpixel can be assumed to have the same class label. The quality of the generated superpixel map greatly depends on the *base image* onto which these algorithms are applied; the *base image* is the result of pre-processing the HSI image before applying segmentation.

Hence, the base image must be chosen carefully, prior to the application of superpixel segmentation algorithms. These algorithms can be applied either on raw HSI or processed (e.g. dimension-reduced) HSI images. When applied to the raw HSI, the very valuable, discriminating spectral information underlying the image is preserved and fully exploited. In [21], an example of segmentation approach addressing the full spectral extent of the original dataset is reported, with the SLIC algorithm. In the general case, however, since HSI may have hundreds of bands, some of which may contribute significant amounts of noise or artifacts, directly segmenting the whole HSIs with the superpixel method will result in very high computational cost and possibly also poor performances due to unsuppressed noisy bands. To address this issue, processed HSIs can be used as the base images for superpixel segmentation, which can greatly reduce the computational cost and improve robustness to noise.

Several types of operations like dimensionality reduction, band selection, or feature extraction can be performed on the original HSI image, to extract only the most informative and significant features from HSI. Dimensionality reduction techniques like PCA can be employed to extract e.g. the first three principal component bands, onto which superpixel segmentation algorithms can be applied. This can significantly reduce the computational complexity and also the impact of noise. Another possible approach is to select the most informative and discriminative bands as the base image upon which superpixel segmentation algorithms are applied instead of deriving new, non-spectral bands. By doing so, the inherent spectral information in HSI can be well preserved while reducing the impact of noise. Feature extraction can be another effective approach to obtain an informative base image. When applied to the extracted features, superpixel segmentation leads to an enhanced spatial structure, thus reducing the effect of incorrect region boundaries, decreasing differences within the same class, and limiting over-segmentation [24]. The Kurtosis Wavelet Energy (KWE) [25] and the Kurtosis Curvelet Energy (KCE) [26] can also be considered as features upon which superpixel segmentation can be applied. Hybrid methods may also be adopted to create the base image. Features may be extracted from the PCA image. After feature extraction, PCA may once again be incorporated on the extracted features to generate the base image [27]. In Figure 3, a scheme is shown, visually summarizing the possible approaches to generating a base image for superpixel formation.

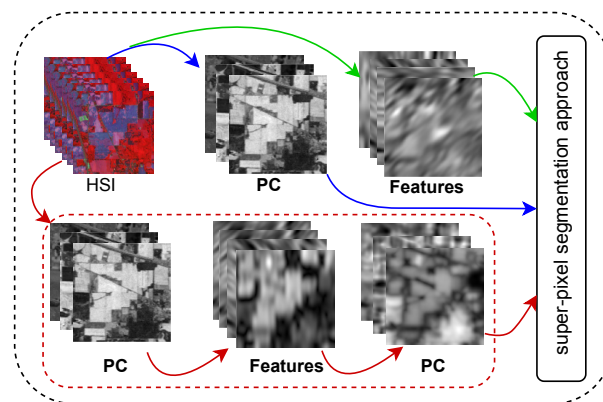


Fig. 3. Base Image for Superpixel Formation. The blue arrows indicate the dimensionally reduced features, the green arrow refers to the Extracted features, and the red arrow refers to the Hybrid features upon which Superpixel Segmentation algorithms are applied.

After obtaining the base image, the next step is application of superpixel segmentation algorithms to generate superpixel maps. A detailed description regarding superpixel creation algorithms is provided in section IV.

Before delving into the various techniques, it is worthwhile to mention a cross-cutting issue, i.e. defining the optimal number of superpixels  $K$ . This is generally a difficult task, still, it is important to solve it opportunely as the quality of the generated superpixel map heavily depends on this factor. Several existing methods use manual approaches to determine a "good" value for  $K$  [10]. However, some automatic methods have also been proposed to determine  $K$  based on the

contextual information in image [27]. If the value of  $K$  is fixed, then a single superpixel map is generated. Often, multi-scale superpixel segmentation maps (Figure. 4) are generated as the use of multiple scales allows the capturing of local spatial structures of various sizes [28]. Instead of varying the value of  $K$ , one may also generate multiple superpixel maps by considering different features (Fig. 4). This approach may very well generate a superpixel map that takes into account all the key features. Such a hybrid multiple-superpixel creation approach has not been deeply explored in literature so far, despite being very appealing.

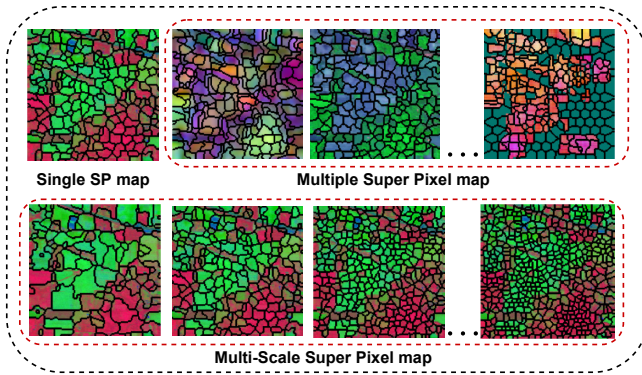


Fig. 4. Characterisation of Superpixel

#### IV. SUPERPIXEL CREATION ALGORITHM FRAMEWORK IN HSI

In recent years, superpixel segmentation algorithms have gained a lot of popularity due to the savings in computational loads that they permit at the subsequent processing stages of HS data. Hence, several segmentation algorithms have been proposed by researchers recently. The existing superpixel segmentation algorithms can be categorized broadly into two groups: Graph-based and Gradient-ascent-based methods (Fig. 5).

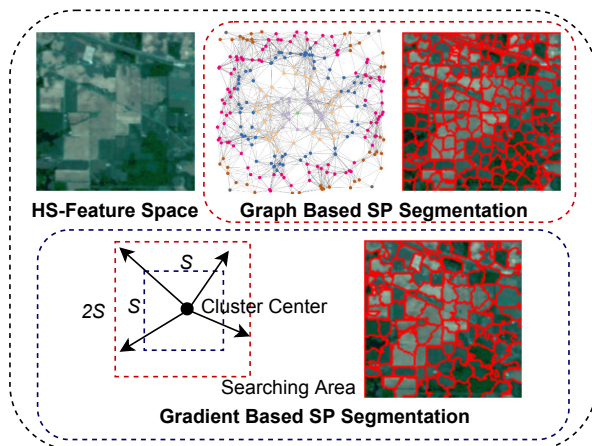


Fig. 5. Superpixel Segmentation Approaches

#### A. Graph-Based Approach

In graph-based approaches, each pixel is treated as a node of the graph and the similarity between neighboring nodes is represented by edge weights. Similar nodes are assigned higher weights. Superpixels are then created by minimizing a cost function defined over the graph [13]. Popular graph-based superpixel segmentation algorithms are: Normalized Cut (NC) [11], Felzenswalb and Huttenlocher (FH) [29], Superpixel Lattice (SL) [30], Constant Intensity Superpixels (CIS) [31], Entropy Rate Superpixels (ERS) [32]. In the field of HSI processing, ERS is the most popular and frequently used graph-based superpixel segmentation algorithm. Instead of applying ERS on the original HSI, it is often applied on the reduced HSI. A detailed description of the ERS algorithm is provided in the subsection below:

1) **ERS:** ERS is a graph-based clustering approach where superpixels are generated by performing graph partitioning [32]. For the generation of superpixel from an image, four steps are followed: Graph construction, entropy rate definition, balancing function definition, and optimization.

- **Graph Construction:** An image is initially mapped to a graph  $\mathcal{G} = (V, E)$ , where  $V$  represents a set of vertices  $v_i$  corresponding to each pixel in an image and  $E$  indicates the corresponding set of edges  $e_{i,j}$  between  $V$ . The edge weight  $w_{i,j}$  measures the similarity between neighboring pixels by calculating the spectral distance. The goal of the superpixel segmentation task is to find  $K$  connected subgraphs from the graph  $\mathcal{G} = (V, E')$  where,  $E' \subseteq E$  is the selected edge set.
- **Entropy Rate:** For the creation of compact and homogeneous superpixels, the criterion used is the entropy rate of the random walk on the constructed graph  $\mathcal{G} = (V, E')$ . The expression for entropy rate function is as follows (Eqn. 1):

$$\mathcal{H}(E') = - \sum_i \mu_i \sum_j P_{i,j}(E') \log(P_{i,j}(E')) \quad (1)$$

where  $P_{i,j}$  is the transition probabilities in  $\mathcal{H}(E')$  which is expressed as follows (Eqn. 2):

$$P_{i,j}(E') = \begin{cases} \frac{w_{i,j}}{w_i} & \text{if } i \neq j \text{ and } e_{i,j} \in E' \\ 0 & \text{if } i \neq j \text{ and } e_{i,j} \notin E' \\ 1 - \frac{\sum_{j:e_{i,j} \in E'} w_{i,j}}{w_i} & \text{if } i = j \end{cases}$$

$$w_i = \sum_{k:e_{i,k} \in E} w_{i,k}, w_T = \sum_{i=1}^{|V|} w_i, \mu_i = \frac{w_i}{w_T} \quad (2)$$

With the inclusion of a new edge, the uncertainty of a jump of the random walk increases. The entropy rate drastically increases when the selected edges form compact and homogeneous clusters.

- **Balancing Function:** To obtain clusters of similar sizes a balancing function is introduced which is defined as in Eqn. 3

$$\mathfrak{B}(E') = \mathcal{H}(Z_{E'}) - \mathcal{N}_{E'} \\ = - \sum_i P Z_{E'}(i) \log(P Z_{E'}(i)) - \mathcal{N}_{E'} \quad (3)$$

where the selected edge set is  $E'$ , the cluster membership distribution is represented by  $Z_{E'}$  and the number of connected components is denoted by  $N_{E'}$ . If  $S_{E'} = S_1, S_2, \dots, S_{N_{E'}}$ , is the graph partitioning for the edge set  $E'$ , then  $Z_{E'}$  distribution is as follows (Eqn. 4):

$$PZ_{E'}(i) = \frac{|S_i|}{|V|}, \quad i = \{1, \dots, N_{E'}\} \quad (4)$$

When edges from similar-sized clusters are selected, the value of the balancing function increases.

- **Optimization:** The objective function for ERS is formulated by combining the entropy rate function and balancing terms together Eqn. 5. Hence, more compact, balanced, and homogeneous clusters can be obtained using ERS. To achieve accurate superpixels, the objective function is optimized on the edge set.

$$\begin{aligned} & \max_{E'} \mathcal{H}(E') + \lambda \mathfrak{B}(E') \\ & \text{subject to } E' \subseteq E \text{ and } N_{E'} \geq K \end{aligned} \quad (5)$$

where the weight of the balancing term is  $\lambda \geq 0$ .

### B. Gradient-Ascent-Based Approaches

In such kind of approaches, an initially defined tentative set of clusters points is iteratively refined using a gradient ascent method until some convergence criteria are met. Some of the popular gradient-based approaches are: Mean Shift (MS) [33], Quick Shift (QS) [34], Watershed [35], Turbopixels (TP) [36], Simple Linear Iterative Clustering (SLIC) [37]. Among these gradient-based approaches, the SLIC algorithm is the most popular superpixel algorithm for HSI processing. A brief description of the SLIC algorithm is presented in the subsection below.

1) *SLIC*: Let the input HSI be denoted as  $H \equiv \{h_1^b, h_2^b, \dots, h_n^b\}$  with  $n$  pixels, where  $\{h_i^b\}$  represents the value at  $i^{th}$  pixel for the  $b^{th}$  spectral band and  $i = 1, 2, \dots, n$ ,  $b = 1, 2, \dots, B$ .  $B$  is the total number of spectral bands. In SLIC, distance is computed within a  $2Q \times 2Q$  window around the cluster center, where  $Q = \sqrt{\frac{n}{K}}$ . The distance between the cluster center and pixel  $i$  is calculated as follows (Eqn. 6):

$$\mathfrak{D} = \mathfrak{D}_{spectral} + \frac{w}{Q} \mathfrak{D}_{spatial} \quad (6)$$

where,  $w$  is the weighting factor between spectral and spatial features. The spectral and spatial distance between pixel  $i$  and  $j$  are represented as in Eqn. 7 and 8

$$\mathfrak{D}_{spectral} = \sqrt{\sum_{i=1}^B (h_i^b - h_j^b)^2} \quad (7)$$

where,  $\mathfrak{D}_{spectral}$  is the measure of homogeneity within the superpixels.

$$\mathfrak{D}_{spatial} = \sqrt{(r_i - r_j)^2 + (u_i - u_j)^2} \quad (8)$$

where,  $(r, u)$  denotes the location of pixel  $i$  in superpixel. The spatial distance  $\mathfrak{D}_{spatial}$  ensures regularity and compactness in the generated superpixels.

### C. A remark

As the reader may note after the above discussion, there exist several superpixel algorithms that are usable in the HSI context. However, it is to be noted that the concept for most of them originated outside HSI. This means that nowadays there exist practically no dedicated superpixel algorithms specifically designed for HSI, and hence capable of leveraging the benefits of HSI data to a full extent. Hence, in our opinion, the development of a dedicated superpixel algorithm for HSI is still an open challenge.

## V. POST-PROCESSING FRAMEWORKS FOR SUPERPIXELS

The generated superpixel map can now serve as a convenient primitive entity, upon which further processing tasks may be carried out to obtain desirable results. As superpixels can adaptively change the object boundaries based upon the available contextual information, they are often used for object-level image analysis. The superpixel may be either used as a guidance map [10] for refining the final classification result or it may be used to directly compute features from it [3]. The different approaches to combining segmentation and classification are treated in the next four subchapters.

### A. Superpixel-Guided Classification

Fig. 6 shows the framework for superpixel-guided classification. An initial classification map is first generated based on the probability description of each pixel belonging to every class. Later, the classification map is further optimized/regularized with the guidance of the segmentation map. Initially, classification maps are obtained by adopting standard feature extraction (Local Binary Pattern (LBP), Gabor, Extended Morphological Attribute Profile (EMAP), etc.) followed by classification (e.g. SVM) approaches. For supervised HSI classification, let there be  $C$  classes present in the scene and  $\mathfrak{J}$  training samples in total for all classes. For each feature,  $t \in T$  ( $t = 1, 2, \dots, T$ ), training set  $X^{train} = [X_1^{train}, X_2^{train}, \dots, X_C^{train}] \in \mathfrak{R}^{B \times T}$  represent the corresponding training samples for each class. The classifier generates  $T$  predicted labels for each pixel using SVM to classify each feature. Finally, the class label of test sample  $y$  is associated with the class having the highest frequency over all  $T$  features, that is (Eqn. 9)

$$Class(y) = \operatorname{argmax}_c \sum_{t=1}^T \delta(\mathcal{P}_t, c), \quad c = 1, 2, \dots, C \quad (9)$$

$$\delta(\mathcal{P}_t, c) = \begin{cases} 1, & \text{if } \mathcal{P}_t = c \\ 0, & \text{if } \mathcal{P}_t \neq c \end{cases} \quad (10)$$

where  $\mathcal{P}_t$  represents the predicted label of the test sample obtained from  $t^{th}$  feature and  $\delta$  is an indicator function (Eqn. 10). Hence, an initial label for each pixel is obtained by the classifier. Simultaneously, superpixel segmentation maps are also computed over the original HSI, hence the location information is the same for both the maps. Let  $S = [S_1, S_2, \dots, S_K]$  denotes the superpixel map, with  $K =$  number of superpixels. In the  $K^{th}$  superpixel, the labels of all the pixels can be

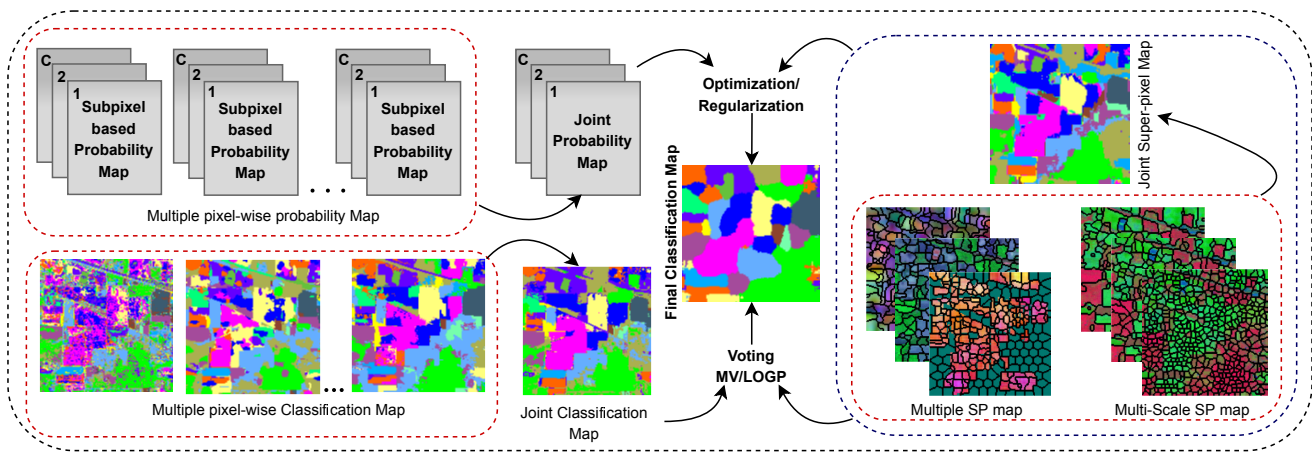


Fig. 6. Fusion multiple guided

identified by adopting the majority voting strategy [38], which is computed by using the following formula (Eqn. 11):

$$Class(S_K) = \underset{c}{\operatorname{argmax}} \sum_{y \in S_k} \delta(Class(y), c), \quad c = 1, 2, \dots, C \quad (11)$$

Hence, the majority voting (MV) strategy (Fig. 7) is mostly utilized to regularize the classification map based on the segmentation map [39]. Majority voting is sub-optimal by design since it only fuses class labels based on a vote over individual class labels from each classifier in the ensemble. Weighted MV (WMV) [40] is an upgraded version of MV, as it weighs the vote of each pixel by the distance between the pixel and its cluster. However, both of them are quite dependent on the preliminary classification accuracy.

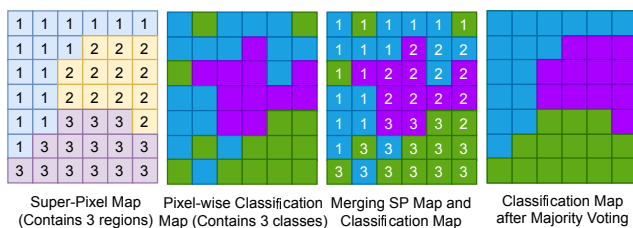


Fig. 7. Majority Voting

Instead of directly assigning class labels for each pixel, it is better to assign a probability description for each pixel belonging to a particular class; this also permits taking into account the presence of mixed pixels in an image. Hence, the class probability of each pixel belonging to each class is determined by using an SVM-based classifier [41]. Let  $P_c(x)$  denote the probability of pixel  $x$  belonging to the  $c^{th}$  class. Then, for each superpixel  $S_i$  in HSI data, the label of all the pixels contained in the superpixel is identified by calculating the following formula (Eqn. 12):

$$Class(S_i) = \underset{c}{\operatorname{argmax}} \sum_{x \in S_i} P_c(x) \quad (12)$$

Hence, decision fusion strategies are applied to classify each pixel based on superpixel-label guidance [10].

Another approach is to perform superpixel-level classification by fusing pixel-level classification outcomes by using the Logarithmic Opinion Pool (LOGP) approach. The main advantage of LOGP is that it treats output from various classifiers independently [20]. The initial class probability obtained from different features can also be combined adaptively by computing joint class probability for each pixel  $x_i$ . The certainty degree of class probability (CDCP) and the confidence score of the classifier (CSC) can be utilized to compute the joint class probability [42]. A higher value of CDCP indicates higher discriminative capability between different classes for a specific pixel. CSC assigns a higher weight to a classifier with higher discriminative ability. The computed joint probability helps in determining the final class estimate for  $x_i$  according to maximum probability. However, some isolated pixels may also exist due to poor probability estimation. Hence, often Markov Random Field (MRF) regularization is adopted to refine the final classification map [43].

### B. Direct Classification using Superpixels

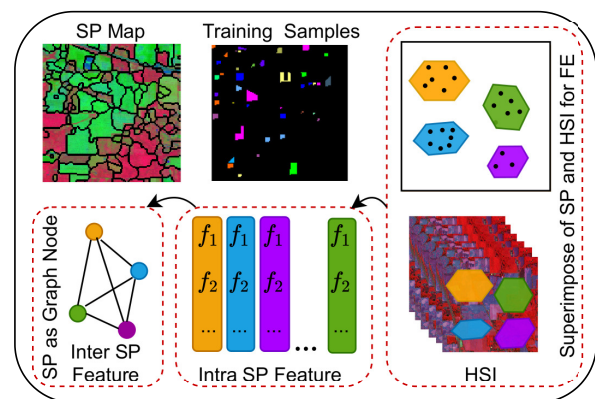


Fig. 8. Direct superpixel classification

Fig. 8, illustrates the framework for direct classification using superpixels. In this approach, features are computed directly upon the superpixels based on which the classification is performed. Superpixels can be used to generate the test set

as they group spatially connected and spectrally similar pixels together into one class. The training set can be constructed by considering labeled pixels from each class. Then, techniques like Convex Hull or Affine Hull models can be utilized to represent training and test sets as these models consider both the variance and similarity within each set to adaptively characterize it. Then set-to-set distance is computed to measure the similarity between train and test set. Finally, a classification label is assigned to each test set based on distance criterion [44], [45].

The generated superpixels are quite informative and hence features can be extracted directly from them. Spatial features within each superpixel can be exploited by using filtering operation  $S_i^{mean}$ . This operation also minimizes the effect of noise in each superpixel. Filters like mean filters, guided filters, non-local filters, Domain Transform Recursive Filter (DTRF), etc. may be used for within-SP feature extraction. The spatial features among superpixels can be exploited by performing a weighted average operation on neighboring superpixels  $S_{(i,j)}$  where,  $j = 1, 2, \dots, J$ .  $J$  represents the number of neighboring superpixels. Since the mean pixel is the representative feature of each superpixel, the weighted average operation can also be applied on the mean pixels  $s_{(i,j)}^{mean}$ ,  $j = 1, \dots, J$  of neighboring superpixels, and a weighted average pixel can be obtained by Eqn. 13

$$s_i^{WA} = \sum_{j=1}^J w_{i,j} \times s_{i,j}^{Mean} \quad (13)$$

where,  $w_{i,j}$  is the weight defined as in Eqn. 14

$$w_{i,j} = \frac{\exp(-\|s_{i,j}^{Mean} - s_i^{Mean}\|_2^2/h)}{\sum_{j=1}^J \exp(-\|s_{i,j}^{Mean} - s_i^{Mean}\|_2^2/h)} \quad (14)$$

where  $h$  is a predefined scalar. Then  $s_i^{WA}$  is assigned to all pixels in each superpixel  $S_i^{SP}$  and all the superpixels constitute a weighted average feature image  $I^{weigh}$  [46]. These extracted features may be further utilized to generate a composite kernel which is used for classification. Hence, superpixels play a crucial role in improving the classifier performance.

Superpixels can also be used to construct undirected graphs  $\mathcal{G} = (V, E, W)$  where each superpixel acts as a graph node. As opposed to traditional graphs, the graphs constructed using superpixels are computationally more efficient. The weight between the two connected superpixels  $S_i$  and  $S_j$  is constructed based on the features extracted from the superpixels [47].

### C. Superpixel-based Sparsity Methods

1) *Sparse-Representation-based Methods*: Let each pixel of HSI be denoted by  $X \in \mathbb{R}^{B \times 1}$ , where  $B$  is the number of spectral bands. The structured dictionary formed by all classes is represented by  $D$ . The sparse representation can be defined for superpixels instead of individual pixels. Let  $S_i^{SP}$  represent a superpixel, which is composed of a number of similar spectral pixels  $[S_{i,1}, S_{i,2}, \dots]$ . As pixels within each superpixel have similar spectral characteristics, their correlation can be exploited by joint sparse regularization [48]. Each superpixel  $S_i^{SP}$  can be appropriately represented by a linear combination

of atoms (training samples from all classes) from dictionary  $D$  as in the following Eqn.15:

$$S_i^{SP} = DA_i^{SP} + N_i^{SP} \quad (15)$$

The sparse coefficient matrix  $A_i^{SP}$  can be calculated as in Eqn. 16

$$\hat{A}_i^{SP} = \operatorname{argmin} \|S_i^{SP} - DA_i^{SP}\|_F \quad s.t. \|A_i^{SP}\| \leq L. \quad (16)$$

where,  $\|\cdot\|_F$  is the Frobenius norm.  $L$  is the sparsity level representing the number of selected atoms in the dictionary. The above optimization problem can be solved by Orthogonal Matching Pursuit (OMP) or Simultaneous OMP (SOMP). After obtaining  $A_i^{SP}$ , the reconstruction residual error is computed. This error is the difference between the original superpixel and the reconstructed superpixel, and it is computed as in the following Eqn. 17:

$$Err(S_i^{SP}) = \|X_i^{SP} - DA_i^{SP}\|_F \quad (17)$$

Finally, the class label of  $S_i^{SP}$  can be obtained through the following Eqn. 18.

$$Class(S_i^{SP}) = \operatorname{argmin} Err(S_i^{SP}) \quad (18)$$

2) *Collaborative-Representation-based Methods*: It has been argued that it is the ‘‘collaborative’’ nature of atoms, as opposed to ‘‘competitive’’, imposed by the sparseness constraint, that actually improves the classification accuracy. Thus, classifiers based on Collaborative Representation (CR) were proposed for HSI classification. The superiority of CR-based hyperspectral classifiers is due to the utilization of the similar training samples from different classes to represent the test pixel. Recently, superpixel-level collaborative representations proved to be much more efficient than pixel-level collaborative representations as they can extract more adaptive, pure, non-overlapping parcels and they utilize convex combinations of pixel sets, thereby resulting in more stable feature representation [49], [50]. To classify unlabelled superpixels  $S$ , a regularized multitask learning model is defined as follows (Eqn. 19):

$$\begin{aligned} \min_{\{\mathfrak{A}, \mathfrak{B}\}} \sum_t \|S^t a^t - D^t b^t\|_2^2 + \lambda \|\mathfrak{A}\|_F + \|\mathfrak{B}\|_2 \\ s.t. \mathfrak{A} = [a^1, \dots, a^T]; \mathfrak{B} = [b^1, \dots, b^T] \\ \sum a_i^t; t = 1, \dots, T \end{aligned} \quad (19)$$

where,  $\{S^t\}_{t=1, \dots, T}$  denotes the convex hull of multiple different  $T$  features extracted from different perspectives of an unlabelled HS superpixel  $S$ .  $\{D^t\}_{t=1, \dots, T}$  represents the corresponding sub-dictionaries of  $S^t$  constructed with features of the same training samples.  $\{a^t\}_{t=1, \dots, T}$  is the pixel weight set and  $\{b^t\}_{t=1, \dots, T}$  is the collaborative coefficient set. The label to be assigned to unlabelled superpixels is determined by minimal total residual (Eqn. 20).

$$Class(S) = \operatorname{argmin}_{i=1, \dots, C} \sum_t \|S^t a^t - D_i^t b_i^t\|_2^2 \quad (20)$$

3) *Low-rank representation approach*: Low-rank representation (LRR) overcomes the limitations of sparse representation by extracting the intrinsic global structure of HSI. Different from the sparse representation, the LRR is no longer sparse on a single pixel, but looks for the common sparsity support of all test samples, which is the lowest rank. Rank is a reasonable measure of the matrix's sparsity [51]. Compared with sparse priors, the low-rank constraint does not need to consider specific correlation patterns and is self-adaptive as well as robust, which can effectively measure the global correlation of the hyperspectral data. The spectral characteristics of neighbor pixels are highly correlated, and the low-rank prior can efficiently preserve the intrinsic structure of the data. Based on the observation that the pixels within a small neighborhood usually consist of similar materials, it is reasonable to enforce the low-rank constraint on their coefficient matrix in a small neighborhood. Hence, superpixels are often employed to extract spatial-structural information by grouping similar pixels together. Let  $S$  be a segmentation map and  $\Sigma = \{r : r = 1, 2, \dots, p\}$  be the region number.  $p$  is the amount of homogeneous regions in  $S$ .  $X = \{x_i \in \mathbb{R}^B, i = 1, \dots, N\}$  is an unknown test set with  $N$  as the number of test samples.  $D \in \mathbb{R}^{B \times J}$  is a structured dictionary consisting of training samples from all classes.  $J$  is the total number of training samples. In a homogeneous region  $S^r$  all pixels make up a matrix  $X^r \in \mathbb{R}^{B \times N^r}$ , where  $N^r$  is the number of pixels in the region  $S^r$ . Then the LRR in each region can be represented as (Eqn. 21):

$$\min_{\alpha} \frac{1}{2} \|X^r - D\alpha^r\|_F^2 + \lambda \|\alpha^r\|_* \quad (21)$$

where  $\alpha \in \mathbb{R}^{J \times N}$  is the coefficient matrix that is to be restored.  $\lambda > 0$  is a scalar regularization parameter. Nuclear norm  $\|\alpha^r\|_*$  is convex relaxation of rank function.

LRR is also widely applied for image denoising applications as it can simultaneously remove different types of noises [52], [18]. HSI is first segmented into homogeneous regions using superpixel segmentation, then a noise model is defined for each region as follows (Eqn. 22):

$$H = H' + E_{sparse} + N_{gaussian} \quad (22)$$

where,  $H$  is the observed HSI and  $H'$  is the clean HSI.  $E_{sparse}$  represents the sparse error term denoting outliers and non-Gaussian noise.  $N_{gaussian}$  is the Gaussian noise present in each homogeneous region. Spectra in each homogeneous region have high correlation, hence have underlying low rank property.  $E_{sparse}$  is expected to be sparse, as the percentage of outliers/non-Gaussian noise  $E_{sparse}$  is much smaller than  $Y$ . Hence,  $E_{sparse}$  has a large number of zero elements. The optimisation problem is formulated as follows (Eqn. 23):

$$\min_{H', E_{sparse}} \|H'\|_* + \lambda \|E_{sparse}\|_1 + \frac{1}{2\mu} \|H - H' - E_{sparse}\|_F \quad (23)$$

$\|H'\|_*$  is the nuclear norm of  $H'$  which is defined as sum of singular values of  $H'$  i.e.  $\|H'\|_* = \sum_{i=1}^r \sigma_i(H')$ .

4) *Non-negative matrix factorization approach*: The non-negative matrix factorization (NMF) approach aims to decompose a given nonnegative matrix  $X^{B \times N}$ , into two smaller

nonnegative matrices  $M \in \mathbb{R}^{B \times S}$  and  $A^{S \times N}$  which can approximately represent matrix  $X$  (Eqn. 24).

$$X \approx MA \quad (24)$$

This non-negativity makes the resulting matrices easier to inspect. The non-negative features helps in learning the spatial structural information [53]. In HSI, NMF is widely used for unmixing applications as it can simultaneously decompose mixed pixels  $X$  into endmembers  $M$  and abundances  $A$ . A cost function (Eqn. 25) is defined to quantify the quality of the approximation, which is constructed according to the distance metrics.

$$\min_{M, A} f(M, A) = \|X - MA\|_F^2 \quad s.t. \quad M \geq 0, A \geq 0 \quad (25)$$

To address the above optimization problem, various algorithms have been proposed. Alternating non-negative least squares (ANLS) is one of the most popular algorithm where optimisation problem is decomposed into two sub-problems [54]. Recently, for effective incorporation of spatial information into the NMF, various sparsity inducing spatial regularizers are introduced. The optimisation problem with spatial group sparsity regularization becomes as follows (Eqn. 26):

$$\min_{M \geq 0, A_r \geq 0} f(M, A_r) = \frac{1}{2} \sum_{p=1}^P \|X^p - MA^p\|_F^2 + \lambda \sum_{p=1}^P \sum_{a_j \in \vartheta_p} \epsilon_j \|W^p a_j\|_2 \quad (26)$$

where,  $A_r = (A^1, \dots, A^p)$  is the abundance matrix divided into  $P$  superpixel groups.  $X^p$  represents pixels belonging to the  $p^{th}$  superpixel.  $A^p = [a_1, \dots, a_{n_p}] \in \mathbb{R}^{S \times n_p}$  is abundance matrix for spatial superpixel group  $\vartheta_p$ .  $\lambda$  controls tradeoff between reconstruction and regularizer.  $\epsilon_j$  is pixel by superpixel confidence index that weights local similarity between pixel and superpixel it belongs to.  $W^p = \text{diag}(w_1^p, \dots, w_S^p)$  is a diagonal matrix that represent superpixel-wise weight matrix.

#### D. Superpixel based Deep Learning

In recent years, deep learning has emerged as a powerful feature extraction tool that can effectively address the existing non-linearity problem in hyperspectral data. As compared to the traditional machine learning algorithms, the deep learning techniques utilize a series of hierarchical layers to extract discriminative features from original data. Texture information and edge information are usually extracted by the initial layers whereas, more complicated features are extracted via the deeper layers. Hence, deep learning algorithms can effectively deal with the large spectral variability of spectral signatures. The most popular deep learning architectures in this context are: stacked autoencoders (SAE) [55], deep belief networks (DBN) [56], Convolution neural networks (CNN) [57], Recurrent Neural Networks (RNN) [58], Generative Adversarial Networks (GAN) [59] etc. These networks are utilized to extract features of HSI. The extracted features may be spectral, spatial, or spatial-spectral. However, instead of extracting either spectral or spatial features, it is better



to extract spatial-spectral features simultaneously. In [60], a SAE-based method for HSI classification is proposed by incorporating a spatial constraint in the energy function to better maintain the spatial information. Dimensionality reduction methods are often combined with CNN for extraction of higher-level spectral-spatial features [61]. For 3D spectral-spatial feature learning, in [57] a 3D CNN is introduced for HSI classification. In [62], a deep pixel-pair feature-based HSI classification method was proposed. Although the HSI classification methods mentioned above have achieved higher-level representations of HSIs, a neighborhood window with a fixed size and shape is used to extract the spatial information without considering the different HSI spatial structures. With a fixed window, the scale information of objects is not taken into account, which results in less structural information and the sample is susceptible to spectral distortion and hence noisy classification result. To overcome these problems, superpixels are incorporated to extract adaptive object information.

In [63], the spectral classification module and the spatial constraints module are employed where the spectral classification module uses a deep network called ‘Stacked Denoising Autoencoders’ (SdA) to learn feature representation of the data. Then pixel-wise classification is performed using a logistic regression model. Finally, the superpixel constraint is used as a spatial constraint to refine the classification result. In [64], the latent relationship between a pixel and the superpixel constraints is integrated into the stacked SAE model. The constraint from superpixels is utilized to avoid the “salt-and-pepper” problem by providing feedback information to the latent relationship learning. The loss functions constructed by superpixels are integrated into the objective function of SAE. To effectively exploit the spatial-spectral information within a superpixel in [65], HSI is first segmented from coarse to fine scales using superpixels. Then, the spatial features within each superpixel and among superpixels are sufficiently exploited by the local and non-local similarity measures. Finally, recurrent neural networks with stacked autoencoders are proposed to learn the high-level multiscale spectral-spatial features. To fully exploit three complementary characteristics of subpixel, pixel, and superpixel a novel HSI feature learning network (HSINet) is developed in [66] which learns consistent features by self-supervision for HSI classification. HSINet contains a three-layer deep neural network and a multifeature convolutional neural network. It automatically extracts the features such as spatial, spectral, color, and the boundary as well as context information. To boost the performance of self-supervised feature learning with likelihood maximization, the conditional random field (CRF) framework is embedded into HSINet. A novel marginal SAE with adaptively spatial regularization (ARMSAE) is proposed in [67] to address the problem of an insufficient training sample problem. Initially, superpixel segmentation is performed. Then, pre-training is performed based on adaptively-spatial regularization to extract contextual information of samples in the homogeneous regions. It utilizes unlabeled adjacent samples to alleviate the lack of training samples. At the fine-tuning stage, the marginal samples based on the geometrical property are selected to tune the ARMSAE network. Finally, the label of each test sample is determined

by all the samples located in the same homogeneous region.

## VI. SUPERPIXELS AND THEIR APPLICATIONS

Due to the underlying properties of superpixels, in recent years it has been successfully applied in a variety of applications like: Classification, Spectral Unmixing, Dimensionality Reduction, Band Selection, Active Learning, Denoising, and Anomaly Detection.

### A. Classification

The prime objective of hyperspectral image classification is to designate a unique class label to each pixel in the image. For accurate classification results, the feature extraction process must be done perfectly. In traditional classification approaches, however, a rectangular window is often utilized as a local probe for extraction of contextual features; this fails to extract local spatial structural information as pixels in a particular area may belong to different classes. Hence, to solve the above problem, superpixel segmentation was introduced into the HSI classification framework as it can adaptively modify its shape and size according to the spatial structural information. The superpixel segmentation map may be used as a guidance map to optimize/regularize the initially obtained classification map to produce the final classification map [10], [39]. Another approach is to compute features directly upon the superpixels based on which classification is performed. Often training and test sets are constructed from superpixels and then the set-to-set distance is computed to measure the similarity between train and test set based on distance criterion [44], [45], [3]. Instead of that, inter- (among) and intra-superpixel (within) features may be computed directly from superpixels. These extracted features are then utilized to generate composite kernels which are then used for classification [46], [42], [24], [27], [9]. Superpixels can also be used as graph nodes to perform graph-based learning for HSI classification [71], [47]. Various sparsity-based classification approaches using superpixels are also proposed in literature to perform HSI classification. Sparse-representation-based methods [81], Collaborative representation based methods [49] and Low-rank representation based methods [66], [51] are some of the popular sparsity based methods used for HSI classification. Various deep-learning-based methods were also proposed in recent years, which utilize superpixels to improve their classification performance for HSI [63], [67], [65], [64].

### B. Spectral Unmixing

In HSI, the problem of mixed pixels is very prominent mainly due to the lower spatial resolution of the HS sensor and atmospheric interference during image acquisition. This results in mixed spectra, in which spectra from different materials may be available on the spectrum of a single pixel. Hence, it is necessary to reduce the effect of mixed pixels as they conceal essential information regarding the pure substance and their distribution in the HSI. The aim of spectral unmixing is to segregate the spectra of the mixed pixel in HSI into a set of constituent spectra (endmembers) and its associated fractional abundances. Endmembers represent the pure

TABLE I  
APPLICATION OF SUPERPIXEL SEGMENTATION IN VARIOUS APPLICATION AREAS FOR HSI

Application	References
<b>Classification</b>	Fang et al. [46], Lu et al. [44], Li et al. [43], Li et al. [28], Lu et al. [42], Jia et al. [68], Liu et al. [24], Liu et al. [69], Yu et al. [70], Jia et al. [10], Fang et al. [27], Cui et al. [71], Jia et al. [39], Cao et al. [45], Sun et al. [9], Mei et al. [72], Duan et al. [73], Chen et al. [74], Jia et al. [75], Jia et al. [76], Ji et al. [77], Hong et al. [78], Sellars et al. [47], Huang et al. [79], Leng et al. [80], Fang et al. [81], Li et al. [49], Zhang et al. [82], Wang et al. [66], Dundar et al. [48], Liu et al. [51], Tu et al. [83], Xu et al. [84], Liu et al. [63], Feng et al. [67], Wang et al. [66], Shi et al. [65], Li et al. [85], Liu et al. [64]
<b>Spectral Unmixing</b>	Wang et al. [14], Wang et al. [86], Borsoi et al. [87], Yang et al. [88], Borsoi et al. [89], Li et al. [90]
<b>Dimensionality Reduction</b>	Mukherjee et al. [91], Jiang et al. [15], Hang et al. [92], Beirami et al. [93]
<b>Band Selection</b>	Yang et al. [16], Tan et al. [94]
<b>Active Learning</b>	Guo et al. [95], Zhou et al. [96], Xue et al. [17], Liu et al. [97], Zheng et al. [98], Zhang et al. [21]
<b>Denosing</b>	Fan et al. [52], Sun et al. [18], Jiang et al. [99], Tu et al. [100]
<b>Anomaly Detection</b>	Ren et al. [19], Huang et al. [101], Huang et al. [102], Gao et al. [103], Huang et al. [104]

substances available in the image, and fractional abundance refers to the percentage of each endmember available in that pixel [105]. The Linear Mixing Model (LMM) is the simplest and most widely used model to represent mixed pixels. Each pixel is modeled as a positive linear combination of all the radiated spectra of the materials making up the pixel. Let,  $X = [x_1, \dots, x_N] \in \mathbb{R}^{B \times N}$  denotes the HSI data matrix with  $B$  bands and  $N$  pixels. Then for the  $j^{th}$  mixed pixel,  $x_j = [x_{1j}, \dots, x_{Bj}]^T \in \mathbb{R}^B$  the LMM representation is represented as:

$$x_j = \sum_{i=1}^S m_i a_{ij} + \varepsilon_j = M a_j + \varepsilon_j \quad (27)$$

$$\implies X \approx MA + \varepsilon$$

where  $S$  denotes the number of endmember signatures.  $M = [m_1, \dots, m_S] \in \mathbb{R}^{B \times S}$  is the mixing matrix consisting of  $S$  distinct endmember spectral signatures.  $a_j = [a_{1j}, \dots, a_{Sj}]^T \in \mathbb{R}^S$  is the abundance vector of  $x_j$ .  $\varepsilon_j \in \mathbb{R}^B$  is the additive noise. In general, the same set of endmembers  $M$  is shared by each pixel in a scene. Hence,  $A = [a_1, \dots, a_N] \in \mathbb{R}^{S \times N}$  and  $\varepsilon = [\varepsilon_1, \dots, \varepsilon_N] \in \mathbb{R}^{B \times N}$ . Here, we need to estimate both  $M$  and  $A$  simultaneously. NMF is the most widely used algorithm for this task. With the incorporation of spatial information to spectral unmixing, significant improvement in both endmember and abundance estimation is observed mainly due to the spatial autocorrelation of the ground surface. Recently superpixels are utilized to incorporate spatial neighborhood information where the shape and size of superpixels are adaptive and are related to the spectral similarity of neighboring pixels. These are utilized to naturally incorporate spatial priors into the unmixing process. In [14], a spatial group-sparsity-regularized NMF is proposed, which incorporates a spatial group sparsity regularizer constraint into the NMF-based unmixing process. Then in [86], a new Group Low-rank constrained NMF technique is developed for linear

HS unmixing. This method combines a low-rank prior of abundances with semantic information. Also, a new NMF method is proposed in [88], which combines non-local spatial information with spatial group sparsity.

### C. Dimensionality Reduction

Dimensionality reduction (DR) is extensively utilized as a preprocessing tool to discard the highly redundant and correlated information in the initial high-dimensional HSI spectral space while at the same time preserving crucial information in a low-dimensional subspace [91]. It can drastically minimize the computational load, and at the same time, effectively train the classifier with a fewer number of available training samples. Several DR techniques have been used in literature, which can be broadly categorized into either supervised or unsupervised methods. Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Minimum Noise Fraction (MNF) are popular linear unsupervised methods. Whereas, locally linear embedding (LLE), neighborhood preserving embedding (NPE), locality-preserving projection (LPP), local pixel NPE (LPNPE), isometric mapping (ISOMAP), and Laplacian eigenmap (LE) are popular non-linear unsupervised methods. The most popular supervised methods are: local Fisher discriminant analysis (LDA), Fisher's linear discriminant analysis (LFDA), generalized discriminant analysis (GDA), nonparametric weighting FE method (NWFE), and decision boundary FE method (DBFE). Generally, these algorithms utilize only spectral features for DR, neglecting the information contained in spatial features. Providing a spatial context to spectral classification algorithms may be helpful because spatially neighboring pixels in images frequently belong to the same classes of interest and, as such, are spectrally similar. In recent years, superpixels came to be widely used for DR applications as a means to incorporate spatial contextual information. In [91], a semi-

supervised feature extraction algorithm is proposed, in which the angular similarity between the spectrally analogous spatial neighbors is minimized, and the angular separation between pixels belonging to different classes is maximized by utilizing unlabeled samples in the projected lower-dimensional subspace. A superpixel-wise PCA approach (SuperPCA) is developed in [15], which applies PCA on each homogeneous region acquired by superpixel segmentation. Further, a multi-scale segmentation-based SuperPCA algorithm (MSuperPCA) is also presented, that can fully exploit the spatial information available in the HSI cube. It effectively integrates the multiscale spatial information to achieve superior classification results by decision fusion. In SuperPCA, however, the individual, traditional PCA implemented in each superpixel will be dominated by those bands with higher spectral variances. In that respect, individual superpixel-based PCA (SuperPCAs) over a homogeneous set of bands can be thought of as a solution. By doing so, SuperPCA can be implemented on only those bands that feature the highest levels of correlation. Accordingly, a method is devised that first categorizes the highly correlated bands of a hyperspectral image via the analysis of its correlation matrix. These band groups are then introduced to a SuperPCA feature extraction [93]. A Spatial Regularized Local Graph Discriminant Embedding (LGDE) approach is used in [92] where a regularization method is proposed to incorporate the spatial information into the LGDE model naturally. Specifically, an over-segmentation method is first used to divide the original HSI into non-overlapping superpixels. Then, for each superpixel, an intra-class graph is constructed to describe the spatial structure information. Finally, the constructed superpixel level intraclass graphs are used as a regularization term for LGDE, resulting in a spatially regularized LGDE (SLGDE). Moreover, to take into account the possible non-linearity of an HSI caused by the complex acquisition process as well as the impacts of atmospheric and geometric distortions, the linear SLGDE model is extended to its kernel counterpart (KSLGDE).

#### D. Band Selection

Band Selection is another approach for dimensionality reduction, in which a smaller subset of the original HS bands is chosen as per a specific criterion. For ranking the spectral bands, either their discriminant capability or their degree of correlation is often considered. For selection of the optimal set of bands, several algorithms have been proposed by researchers that only utilize spectral features, neglecting spatial information. In HSI, the neighboring pixels of the original land surface has a higher probability of belonging to the same class as these have a spatial correlation. Hence, it is more likely that the adjacent similar pixels belong to the same class; this information should be intercepted by superpixels.

Different from the rigid structure of the pixel grid, the boundaries of superpixels align well with the natural object boundaries. An unsupervised BS approach proposed in [16] considers both the spatial and spectral information for accurate HSI classification. First, the ERS algorithm is employed to construct several smaller spectral homogeneous and spatial neighboring pixel chunklets called Superpixel Chunklets

(SCs). Based on the observation that the produced SCs achieve higher homogeneity and consistency within land-cover classes, two bands criteria (BCs) are defined: Metric Learning-based BC (ML-BC) and Representation Learning-based BC (RL), by estimating the optimal transformation through the relevant component analysis (RCA). In the ML-based scheme, to assess the within-SC covariance, the whitening transformation of RCA is utilized. In the RL-based strategy, the discrimination capability of the individual bands is determined by exploiting the within-SC and the total variability. Next, the learned BC is provided as an input to the Affinity Propagation (AP) algorithm. Finally, highly discriminative and weakly-redundant band subsets are selected by using the AP algorithm. Similar to [16], an HS Band Selection strategy was developed for Lithologic Discrimination and Geological Mapping in [94]. A new spectral-spatial structure, i.e. the lithologic superpixel (LS), is constructed by using an improved SLIC superpixel algorithm based on spectral angle distance (SAD), which uses spatial correlation combined with spectral information.

#### E. Active Learning

Accurate classification of HSI requires sufficient high-quality informative training data. During HS data collection, however, only the spectral information for each pixel is acquired by an instrument, whereas the label information is often manually acquired by experts. Hence, the label annotation process is often time-consuming, tedious, and expensive. This results in the availability of very few initial labeled samples available for training. To overcome this, Active Learning (AL) is adopted where the labeling process is guided by certain defined rules instead of the subjective opinion of the human interpreter who tags samples with labels. Active learning is quite popular in the remote sensing community, and a detailed survey on these techniques is provided in [106], [107]. Recently, a lot of attention is gained by those AL approaches which use the joint spectral-spatial information in HSI. Based on the spatial information incorporation process, the existing AL strategies can be broadly categorized into three types: Postprocessing, Preprocessing, and Integrated approaches. In preprocessing-based methods, the spatial-spectral features are directly fed as an input to the AL systems. In the case of postprocessing-based approaches, the regularization of the final classification map is performed by using the spatial segmentation results. Integrated methods, on the other hand, guide the query selection by exploiting the spatial information during the AL process. Depending on the query strategy, AL strategies can be classified into Uncertainty sampling or Breaking Tie [107]. The spatial-contextual information generated by superpixels is found to be quite effective in enhancing the AL process [97]. In [95], gray-level co-occurrence matrix (GLCM) based texture features are first derived from superpixels and are integrated into an AL framework. Next, at each iteration, unlabelled samples are classified using the subspace-based multinomial logistic regression (MLRsub). Then to select the most informative samples, the breaking ties (BT) criterion is utilized. A new approach combining AL and semi-supervised learning (SSL) for HSI classification is proposed in [96].

Initially, morphological component analysis (MCA) is adopted for the decomposition of the original HSI into morphological components. Next, the enlargement of the training dataset based on superpixels is carried out by combining AL and SSL in each feature domain. At last, the decision fusion is performed for the integration of the predictions from extracted components. Spatial uncertainty and spatial homogeneity must also be considered to improve the AL performance. In [17], an enhanced uncertainty measure considering the neighborhood information is proposed. The SLIC algorithm is utilized for generating superpixels, where the selected batch samples are constrained to be from different superpixels, in order to improve the diversity of the selected samples. A superpixel-guided training sample enlargement strategy is developed in [98] to deal with the problem of the small size of the training sample set. First, superpixel segmentation is performed and only those superpixels that contain training samples belonging to no more than one class are explored, and all the pixels of each of these superpixels are assigned to the class of the training samples it contains. Next, with the identified labels, all these classified pixels are added to the initial training sample set for training sample set enlargement. Later, using this enlarged training sample set, the distance-weighted linear regression classifier (DWLRC) is applied to classify each mean vector of each SP. Finally, the last classification map is obtained by assigning each SP with the same label as its mean vector. In [21], the learned superpixel map and initial classification maps are utilized to select the pseudo-labeled samples (PLSs). For the superpixel which contains labeled training samples, the labels of the pixels in this superpixel are likely to be the same as that of the labeled training sample. Hence, the label of pixels in that superpixel made same as that of the labeled training sample. For the superpixels which do not contain labeled training samples, another condition is considered: if the labels of all pixels in that superpixel are the same, this is assumed to be an indicator that the superpixel has sufficient local homogeneity, and these pixels are also selected as PLSs.

#### F. Denoising

The obtained HSI is usually corrupted with various types of noise, which not only scale down the visual quality of HSI, but especially has an adverse effect on the subsequent image processing steps. For several HSI applications like spectral unmixing, super-resolution, object classification, and matching, HSI denoising act as a crucial preprocessing step. The spatial-spectral information should be simultaneously considered to suppress noise. Total variation (TV) denoising [108], tensor decomposition-based denoising [109], multi-dimensional wavelet packet transform (MWPT) [110], and sparse-representation-based denoising [111] are some of the popular HSI denoising techniques recently proposed by researchers. Due to the lack of prior knowledge, however, the aforementioned methods cannot eliminate more than two types of noise. However, for the real-world HSIs, the primary source of noise is not only the additive noise. Various other types of noise like gaussian, impulse, and stripes noise also exist

in practice. To remove these mixed noise, low-rank representation (LRR)-based approaches are often used [52], [18]. In [52], superpixel segmentation is integrated into the LRR based method. LRR helps in the removal of different types of noise simultaneously. Meanwhile, superpixel segmentation is utilized to extract the spatial information of HSI, which can further boost the performance of LRR-based denoising. A novel, fast superpixel-based subspace low-rank learning method is presented in [18] which explores the spatial low rankness within superpixel-based regions for HSI denoising. The method simultaneously imposes the spatial correlation and spectral low-rank properties of the HSI.

#### G. Anomaly Detection

Anomaly Detection is an unsupervised target detection technique, which detects a specific target or anomaly against a complex background without any prior information. The term *anomalies* refers to pixels that have distinct spectral-spatial differences with respect to their surroundings, together with a lower probability of appearance.

For the detection of anomalies, specialized detectors have been developed, which utilize the difference between the background pixels and anomaly pixels. The existing anomaly detection approaches can be broadly divided into two categories: Statistical modeling-based and geometric modeling-based techniques. The statistical modeling-based approaches primarily concentrate on the spectral feature differences and are generally correlated with the Gaussian distribution. The most popular statistical modeling-based method is Reed-Xiaoli (RX) technique. It usually employs multivariate Gaussian distributions for modeling the background. In geometric modeling-based approaches, the main focus is on the difference in spatial distribution. These techniques usually assume that a group of primary spectra can be utilized for the approximate construction of the background pixels. But, it cannot be employed for the representation of the sparse anomaly targets. The sparse representation-based, collaborative representation-based, and low-rank representation-based detectors have been developed recently for anomaly detection. Also, neural networks with hybrid algorithm of CNN and multilayer perceptron (CNN-MLP) can be suggested for anomaly/target detection [112]

In [19], for anomaly detection, a superpixel-based dual-window (SPDW) RX is proposed. In the case of the conventional dual window, both the outer and inner windows are rectangular and have fixed sizes. Whereas for the SPDW, only the outer window is rectangular and has a variable size, but the inner window has an irregular shape. Another approach in [101] considers the similarity between anomaly pixels, along with the variation between anomaly pixels and background pixels for detection. The approach consists of three key steps. First, anomaly queries are automatically generated by applying RX for measuring the spectral difference between background pixels and anomaly pixels. Next, for the characterization of spatial similarity between adjoining nodes (each node refers to a superpixel), a closed-loop graph is constructed. At last, the manifold ranking technique is adopted to assign a ranking

value to each node. Lastly, a final detection result is generated by normalizing the ranking value of each node. In [102], for the detection of anomalies, a game theory-based detection approach is proposed for HSI. In these approaches, for each superpixel, one specific payoff function is defined. The payoff function consists of several energy functions, which can jointly exploit the spectral-spatial features of anomalies and similarities among superpixels. By utilizing multiple energy functions, the detector can be made more powerful in a complex background.

## VII. EXPERIMENTAL ANALYSIS OF EXISTING SUPERPIXEL SEGMENTATION APPROACHES

For experimental analysis of popular superpixel segmentation algorithms three benchmark datasets, namely: Indian Pines, Pavia University, and Houston 2013 were used. Subsection VII-A contains a concise explanation of these datasets. The description of various compared state-of-the-art superpixel segmentation approaches is presented in subsection VII-B. The result and discussion are presented in subsection VII-C.

### A. Dataset Description

1) *Indian Pines*: The dataset was acquired over the North-western Indiana region by using the AVIRIS sensor. It features 220 spectral bands, with wavelengths between 0.4 to 2.5  $\mu\text{m}$ . About two-thirds of the imaged area consists of agricultural land, and the rest contains forests. It is a quite challenging dataset due to the presence of highly mixed pixels due to the low spatial resolution (20 m/pixel) of the sensor. Also, there exists a severe mismatch in the number of samples collected per class that further complicates the classification task. The scene contains sixteen classes and has a size of  $145 \times 145$  pixels. Figure 9 contains the pseudo color image, GT map and class names for the dataset.

2) *Pavia University*: This dataset was collected over the University of Pavia, Italy by using the ROSIS sensor. It originally contains 115, spectral bands in the wavelength range of 0.43 to 0.86  $\mu\text{m}$ , and it has a spatial resolution of 1.3 m. After the removal of noisy channels, the remaining 103 bands are used for our analysis. The image size is  $610 \times 340$  pixels, and the GT is composed of nine challenging classes with similar spectral reflectance trends. The false-color image, GT, and class names are provided in Figure. 10.

3) *Houston 2013 Dataset*: The data was captured using the ITRES CASI-1500 sensor, over the campus of the University of Houston and the neighboring areas of it in Texas, USA. The "IEEE GRSS Data Fusion Contest 2013" also utilized this dataset. The image size is  $349 \times 1905$  pixels and has a spatial resolution of 2.5 m. 144 spectral channels are available in this dataset in the wavelength range of 364 to 1046 nm. In the scene, 15 challenging classes are defined. Figure 11 contains the pseudo-color image, GT and class names for the dataset.

### B. State-of-the-art superpixel segmentation approaches for HSI classification

A comparison of various benchmark state-of-the-art superpixel segmentation algorithms for HSI classification is presented in this section.

1) *SCMK [46]*: It stands for superpixel-based classification via multiple kernels (SCMK). In this approach, the spatial-spectral information within and among the superpixels is utilized via multiple kernels to enhance the classifier's performance.

2) *ULBP-SPG [10]*: a uniform, local, binary-pattern-based superpixel guidance (ULBP-SPG) approach for HSI classification, which first extracts local image features by employing ULBP and later refines the classification map with the guidance of a superpixel map.

3) *R2MK [69]*: is a region-based relaxed multiple kernel(R2MK) technique that fuses multiscale spatial features and spectral features via a kernel collaborative representation classification approach.

4) *ASGSSK [9]*: it is an adjacent superpixel-based generalized spatial-spectral kernel (ASGSSK) method that can very well preserve the image structure by introducing the AS approach to GSSK.

5) *ASMGSSK*: a multiscale framework is adopted here on the ASGSSK method to achieve superior classification performance and solve the problem of optimal superpixel number selection.

6) *SuperPCA [15]*: is a superpixel-based PCA approach that can effectively learn the intrinsic low-dimensional features in HSI even in noisy conditions.

7) *MsuperPCA*: is a multiscale segmentation-based SuperPCA model (MSuperPCA) which can effectively integrate multiscale spatial information to obtain the optimal classification result by decision fusion.

8) *IAP [78]*: it extracts the spatial invariant features by exploiting isotropic filter banks or convolutional kernels on HSI and spatial aggregation techniques (e.g., superpixel segmentation) in the Cartesian coordinate system.

### C. Result and Discussion

Classification results for the Indian Pines dataset with 3% training samples from each class are provided in Table II. Figure 12 contains classification maps for the different superpixel segmentation methods that were compared. From the classification results, it can be clearly noticed that approaches adopting the multiscale segmentation strategy achieve better classification performances, because they consider different scale structures of the particular scene. Also, it solves the problem of selecting the optimal superpixel scale. R2MK, ASMGSSK, and MsuperPCA approaches show better classification accuracy levels as compared to their single-scale segmentation counterparts. Still, the performance of the ASMGSSK technique is superior amongst all other approaches. From the classification map plot in Figure 12. it can also be clearly noticed that the map for the ASMGSSK approach closely resembles the ground truth map.

The second analysis was conducted on the Pavia University dataset. In Table III, the classification result with 30 training samples from each class is provided and the corresponding classification map is displayed in Figure. 13. The performances of all superpixel segmentation algorithms are comparable. Yet, in the case of the ASMGSSK algorithm, the highest improvement of about 3.09 % in OA is observed as compared to

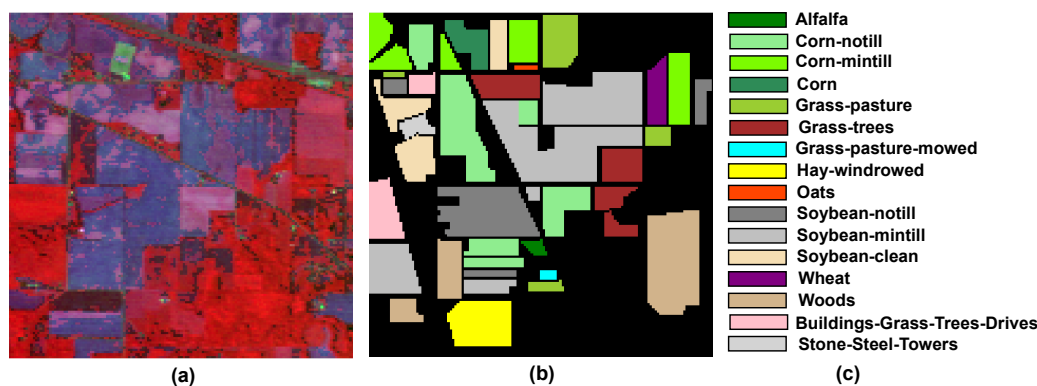


Fig. 9. (a) False Color Composite Image, (b) Ground Truth Image and (c) Class names for Indian Pines Dataset

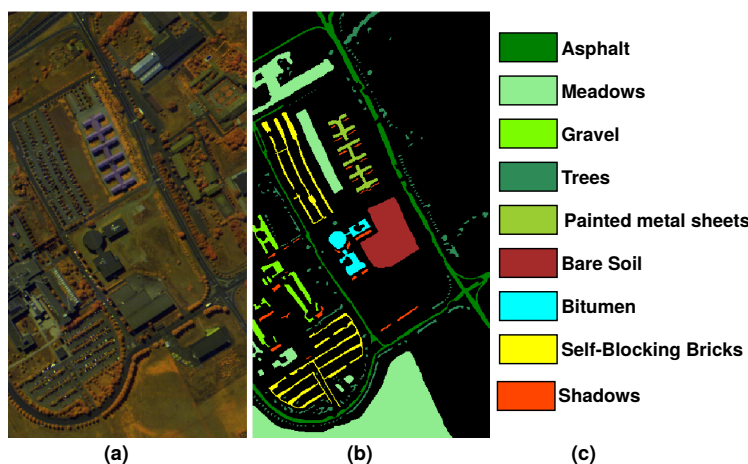


Fig. 10. (a) False Color Composite Image, (b) Ground Truth Image and (c) Class names for Pavia University Dataset

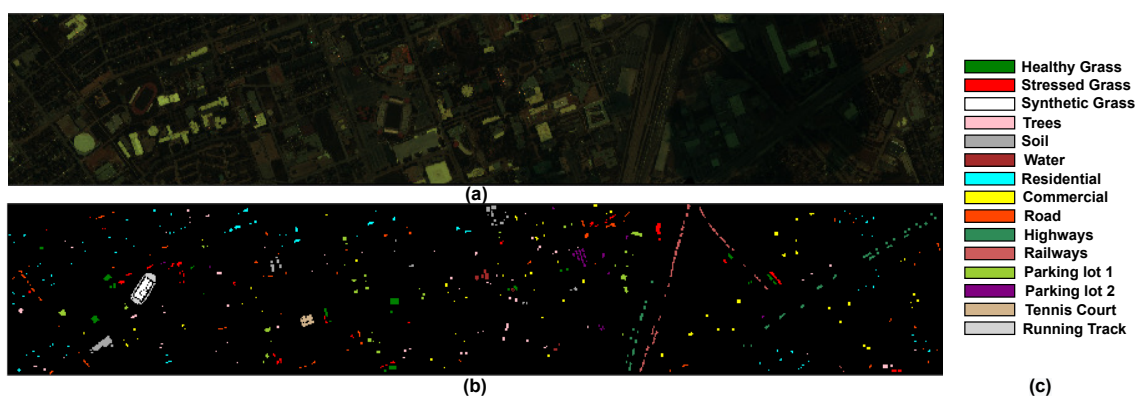


Fig. 11. (a) False Color Composite Image, (b) Ground Truth Image and (c) Class names for Houston-2013 Dataset

SCMK approach. The next rank goes to the IAP algorithm with an increment of about 2.23 % in accuracy. The performance of R2MK, ASGSSK and MsuperPCA are almost identical. ULBP-SPG and SuperPCA classification results are very much similar to SCMK algorithm. Similar conclusions can be drawn from the classification map also (Figure 13)

The third experiment was conducted on the Houston 2013

dataset. To evaluate the performance of the investigated approaches, 50 labeled samples from each class were selected as training samples. In Table IV, the classification results are presented. A visual comparison of performance among the different methods is displayed in Figure 14. For this dataset also, ASMGSSK and IAP shows the best classification results as compared to other tested algorithms. The performance of

TABLE II  
CLASSIFICATION RESULT FOR INDIAN PINES DATASET WITH 3% TRAINING FOR SCMK, ULBP-SPG, R2MK, ASGSSK, ASMGSSK, SUPERPCA, MSUPERPCA, AND IAP ALGORITHMS

#	Samples	SCMK	ULBP-SPG	R2MK	ASGSSK	ASMGSSK	SuperPCA	MsuperPCA	IAP
1	46	95.65	100	95.65	100	100	100	95.65	97.83
2	1428	94.26	96.43	95.87	95.52	96.5	93.56	96.08	96.15
3	830	89.76	91.2	94.58	93.01	94.7	89.71	94.58	94.58
4	237	99.58	94.94	99.58	98.31	96.2	89.89	99.58	99.58
5	483	94	91.93	95.86	95.24	95.86	98.34	93.79	94
6	730	97.53	96.16	98.08	97.53	97.53	99.45	97.95	97.81
7	28	82.14	92.86	92.86	92.86	92.86	100	92.86	92.86
8	478	99.58	99.37	98.54	98.33	99.58	99.72	99.37	99.37
9	20	85	70	60	90	70	46.67	70	65
10	972	90.53	91.46	96.09	95.37	95.58	93.96	94.03	95.68
11	2455	96.05	97.52	98.21	98.37	99.31	95.06	98.21	98.53
12	593	86.34	92.58	91.74	91.4	95.78	84.04	93.09	94.94
13	205	88.78	97.07	99.02	97.07	98.54	100	99.02	96.59
14	1265	98.34	97.79	98.58	97.39	98.97	97.05	99.29	99.05
15	386	99.48	94.3	97.93	95.85	99.22	93.45	95.08	99.48
16	93	95.7	95.7	95.7	95.7	95.7	98.55	95.7	95.7
<b>OA::</b>		94.66	95.49	96.87	96.35	97.47	94.5	96.71	97.15
<b>AA::</b>		93.3	93.71	94.27	95.75	95.4	92.47	94.64	94.82
<b>Kappa::</b>		93.92	94.86	96.43	95.84	97.12	93.73	96.25	96.75

TABLE III  
CLASSIFICATION RESULT FOR PAVIA UNIVERSITY DATASET WITH 30 TRAINING SAMPLES FOR SCMK, ULBP-SPG, R2MK, ASGSSK, ASMGSSK, SUPERPCA, MSUPERPCA, AND IAP ALGORITHMS

#	Samples	SCMK	ULBP-SPG	R2MK	ASGSSK	ASMGSSK	SuperPCA	MsuperPCA	IAP
1	6631	94.5368	95.1	96.2771	97.5809	98.736	95.691	95.42	98.475
2	18649	99.6991	99.5	99.7598	99.7822	99.97	99.8085	99.73	99.935
3	2099	75.5955	84.87	85.7072	83.5732	90.83	80.6028	87.04	86.0176
4	3064	89.7769	93.64	96.1578	96.0218	98.53	90.8394	95.13	96.6998
5	1345	100	100	99.4578	100	100	100	99.9	100
6	5029	96.1453	96.42	98.2394	98.3844	99.7	96.3474	96.82	99.0274
7	1330	88.7951	81.11	91.0658	92.4765	96.15	92.1719	85.46	94.6034
8	3682	93.0155	92.09	91.9072	93.5201	95.99	96.8902	92.11	92.6505
9	947	95.7974	98.81	96.1496	96.4835	97.89	93.2185	99.16	96.789
<b>OA::</b>		95.6	96.11	97.06	97.37	98.69	95.69	96.66	97.83
<b>AA::</b>		92.6	93.5	94.97	95.31	97.53	92.84	94.53	96.02
<b>Kappa::</b>		94.14	94.83	96.15	96.52	98.26	94.27	95.56	97.12

R2MK, ASGSSK and MsuperPCA algorithms are similar. SuperPCA, ULBP-SPG and SCMK results are identical for this case also.

## VIII. OPEN CHALLENGES

### A. Determination of number of superpixels

Determination of the number of superpixels is an important function because this factor controls the segmentation scale. It is a quite challenging task to select an optimal value of  $K$  for the inexperienced users. Depending on the image structure, the correct value of  $K$  must be chosen. Usually, for a dataset having more complex structure and texture information, a larger value of  $K$  must be selected, whereas for a homogeneous dataset the value of  $K$  must be smaller. A larger number of superpixels implies a relatively smaller number of pixels in each superpixel. In that case, the same object may end up being separated into various superpixels. This may result in an inferior classification performance under a small training sample condition. On the contrary, when the number of superpixels is too small, it implies a relatively larger number of pixels in each superpixel. Also, within a single superpixel, several objects may be included. Hence, incorrect classification results may be obtained.

In most cases, the value of  $K$  is chosen manually ([20], [24], [39], [113], [21]) based on observation and experience of the user. In very few works, automatic estimation of the correct  $K$  value is done. In [46] and [27], a texture ratio is computed to characterize the level of complexity of the texture in HSI. Based on the value of such ratio, the value of  $K$  is determined. Even edge detection can be introduced for estimating  $K$  [44], [73]. Complex structures generally include more edges and hence, more superpixels are expected. Still, these methods for automatic estimation of  $K$  are prone to errors as it is practically difficult to determine an optimal value for the number of superpixels that can adapt to all materials. Hence, the multiscale superpixel segmentation approaches were proposed. Different objects exhibit indiscriminately smaller or larger portions in the spatial domain; hence, it is justifiable to fuse the multiscale superpixel segmentation maps so that the spatial structure of various objects can be exploited in a unified framework [76]. By doing so, one can overcome the deficiencies of each scale to spatial constraints and effectively improve the classification performance [114]. Multiscale segmentation approaches, however, increase the computational complexity. Hence, there is a need for development of more advanced strategies for optimal estimation of number of superpixels.

TABLE IV  
CLASSIFICATION RESULT FOR HOUSTON 2013 DATASET WITH 50 TRAINING SAMPLES FOR SCMCK, ULBP-SPG, R2MK, ASGSSK, ASMGSSK, SUPERPCA, MSUPERPCA, AND IAP ALGORITHMS

#	Samples	SCMK	ULBP-SPG	R2MK	ASGSSK	ASMGSSK	SuperPCA	MsuperPCA	IAP
1	1073	96.37	98.59	96.51	98.37	98.43	98.25	97.36	96.96
2	810	96.79	96.35	93.78	97.77	97.75	98.01	97.54	98.87
3	697	96.19	92.91	99.38	95.99	98.6	99.23	95.43	97.66
4	1053	94.77	97.66	94.23	98.79	95.2	94.73	95.47	98.86
5	1242	99.08	98.91	99.75	99.74	99.91	99.75	99.74	99.82
6	325	85.61	91.7	95.67	82.78	80	95.67	89.18	90.33
7	978	81.03	85.33	87.62	88.47	89.45	85.98	87.92	88.77
8	624	77.49	82.53	86.12	85.12	90.04	86.79	91.03	85.93
9	1031	77.72	75.91	79.82	89.44	89.83	83.39	86.49	89.05
10	382	91.66	93.47	89.31	91.59	94.86	89.82	91.76	97.7
11	114	83.93	78.6	85.85	84.86	87.24	85.09	86.91	83.73
12	1233	87.18	94.43	93.92	95.2	97.09	85.65	93.7	96.12
13	449	88.11	84.56	88.12	77.75	77.96	80.05	72.05	75.64
14	428	98.71	99.47	99.21	95.23	98.48	96.32	95.27	91.37
15	660	98.39	99.51	99.35	99.67	98.03	98.69	97.91	99.01
<b>OA::</b>		89.56	90.8	91.7	92.74	93.69	91.4	92.48	93.31
<b>AA::</b>		90.2	91.33	92.58	92.05	92.86	91.83	91.85	92.66
<b>Kappa::</b>		88.72	90.06	91.02	92.16	93.18	90.7	91.87	92.77

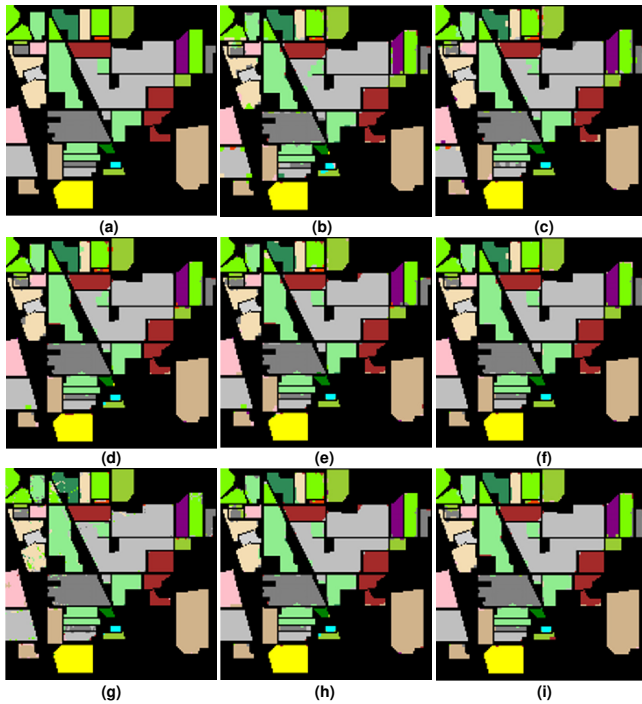


Fig. 12. (a) Ground Truth Image, Classification Maps of (b) SCMCK (c) ULBP-SPG (d) R2MK (e) ASGSSK (f) ASMGSSK (g) SuperPCA (h) MsuperPCA (i) IAP for Indian Pines dataset

### B. Position of superpixel seed points

The initial cluster seed points in the superpixel segmentation algorithms must be carefully chosen in order to obtain accurate segmentation maps. In the case of a standard SLIC algorithm, square grids are used for generating the initial cluster seed points [13]. Hexagonal grids are proposed in place of square grids in [14] as each corner and each side of the hexagon is shared by three and two hexagons, respectively. Compared with the original SLIC based on a square grid, choosing a hexagonal grid for image segmentation has noticeable benefits as it can adequately learn the neighboring spatial information. More non-diagonal neighbors are present for each hexagon

rather than a square. Also, hexagonal grids generate less distance distortion of boundary pixels [77].

In [73], the position indexes of pixels within the  $i^{th}$  superpixel map  $S_i$  are employed on the  $i^{th}$  edge-preserving feature  $F_i$  to extract the corresponding non-overlapping superpixels. Hence, superpixels are generated within the edge-preserving feature.

### C. Superpixel evaluation metrics

The standard superpixel evaluation metrics developed for RGB images are: Boundary Recall, Under-segmentation error, Explained Variation, and Compactness. Let  $S = \{S_j\}_{j=1}^K$  and  $G = \{G_i\}$  represents the superpixel segmentation and GT segmentation respectively for the same image  $I$ . A brief explanation of the aforementioned matrices is provided below.

- **Boundary Recall (Rec):** It computes the fraction of ground truth edges that fall within at least two pixels of a superpixel boundary. A higher value of boundary recall indicates that few number of true edges are missed and hence, better boundary adherence [115]. Rec is expressed using Eqn. 28

$$Rec(G, S) = \frac{TP(G, S)}{TP(G, S) + FN(G, S)} \quad (28)$$

where,  $FN(G, S)$  and  $TP(G, S)$  are the number of false negative and true positive boundary pixels in  $S$  with respect to  $G$  respectively.

- **Under-segmentation error (UE)** It computes the error made by the algorithm while segmenting an image with respect to the GT [37]. Let  $g_1, g_2, \dots, g_Y$  be the GT segments and  $s_1, s_2, \dots, s_K$  represents the superpixel output. Then, the UE error for GT segment  $g_i$  can be expressed using Eqn. 29

$$UE = \frac{1}{F} \left[ \sum_{i=1}^Y \left( \sum_{s_j | s_j \cap g_i > O} |s_j| \right) - F \right] \quad (29)$$

where,  $|\cdot|$  denotes the size of segments in pixels and  $F$  is the image size in pixels.  $O$  is the minimum number of



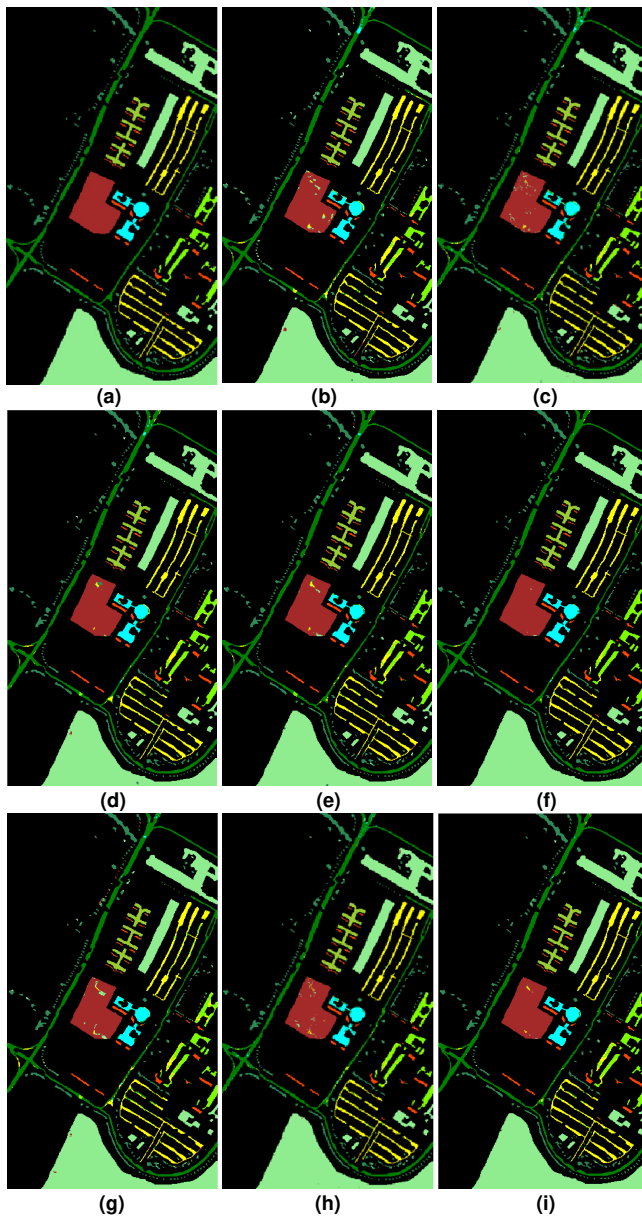


Fig. 13. (a) Ground Truth Image, Classification Maps of (b) SCMK (c) ULBP-SPG (d) R2MK (e) ASGSSK (f) ASMGSSK (g) SuperPCA (h) MsuperPCA (i) IAP for Pavia University dataset

pixels required for overlapping.  $s_j \cap g_i$  is the overlap error of superpixel  $s_j$  with respect to ground truth segment  $g_i$ . The value of  $O$  is set to 5% of  $|s_j|$

- **Explained Variance (EV):** It is used to quantify the variation in the image based on superpixel segmentation quality without depending on the ground truth. It mainly assesses the strong changes in color and structure exhibited by the image boundaries [30]. EV is expressed as in Eqn. 30.

$$EV(S) = \frac{\sum_{s_j} |S_j| (\mu(S_j) - \mu(I))^2}{\sum_{x_n} (I(x_n) - \mu(I))^2} \quad (30)$$

Where  $\mu(S_j)$  is the mean color of the superpixel  $S_j$  and  $\mu(I)$  is the mean color of image  $I$

- **Compactness (CO) :** It compares each superpixel ( $S_j$ ) area ( $Ar(S_j)$ ) with the area of a circle which is having the same perimeter ( $Peri(S_j)$ ) (Eqn. 31) [116]. The higher the value of  $CO$  the better it is.

$$CO(G, S) = \frac{1}{F} \sum_{S_j} |S_j| \frac{4\pi Ar(S_j)}{Peri(S_j)} \quad (31)$$

Even though the above metrics are popular for natural RGB images, for hyperspectral images it is difficult to compute these parameters. Hence, to assess the performance of the segmentation algorithm, the standard metrics generally used for evaluating the classification performances (overall accuracy (OA), average accuracy, and kappa coefficient ( $\kappa$ )) are adopted. If for a particular superpixel segmentation algorithm, a substantial enhancement in classification performance is observed, then it means that a particular segmentation algorithm is superior as compared to other algorithms. At the moment, however, no dedicated segmentation evaluation metrics seem to have been developed yet for hyperspectral images. Hence, more work needs to be done in this regard.

#### D. Selecting the base image for generation of superpixels

The base image upon which the superpixel segmentation algorithm is applied impacts heavily on the final segmentation result. Hence, the choice of the most suitable base image before applying the superpixel segmentation algorithm is a crucial decision. These algorithms may be applied directly to the raw HSI or processed (dimensionality-reduced, band-selected, or feature-extracted) HSI.

#### E. Selecting an appropriate superpixel segmentation algorithm for HS image

Several superpixel segmentation algorithms are available in literature which were originally developed for computer vision applications. These algorithms basically expect a natural RGB image as input and were conceived under this assumption. While applying superpixel segmentation algorithms, often the dimension of HS images are down-scaled in accordance with the segmentation algorithm. However, there exist no dedicated superpixel segmentation algorithms designed for HSI, despite these could enjoy the full benefits of HSI data and its information content. Hence, finding a dedicated superpixel algorithm for HSI is still an open challenge. Moreover, HS compression techniques can be considered when generating superpixels.

## IX. CONCLUSION AND FUTURE SCOPE

In recent years, superpixels have been extensively applied in a broad range of applications due to their inherent properties. Superpixels can partition a particular image into several smaller meaningful regions upon which features can be computed. Hence, by using superpixels the computational complexities can be reduced drastically as superpixel-based processing significantly reduces the input entries for the subsequent algorithms. Due to the aforementioned benefits, superpixels have been widely employed in various application areas of hyperspectral image processing like classification,

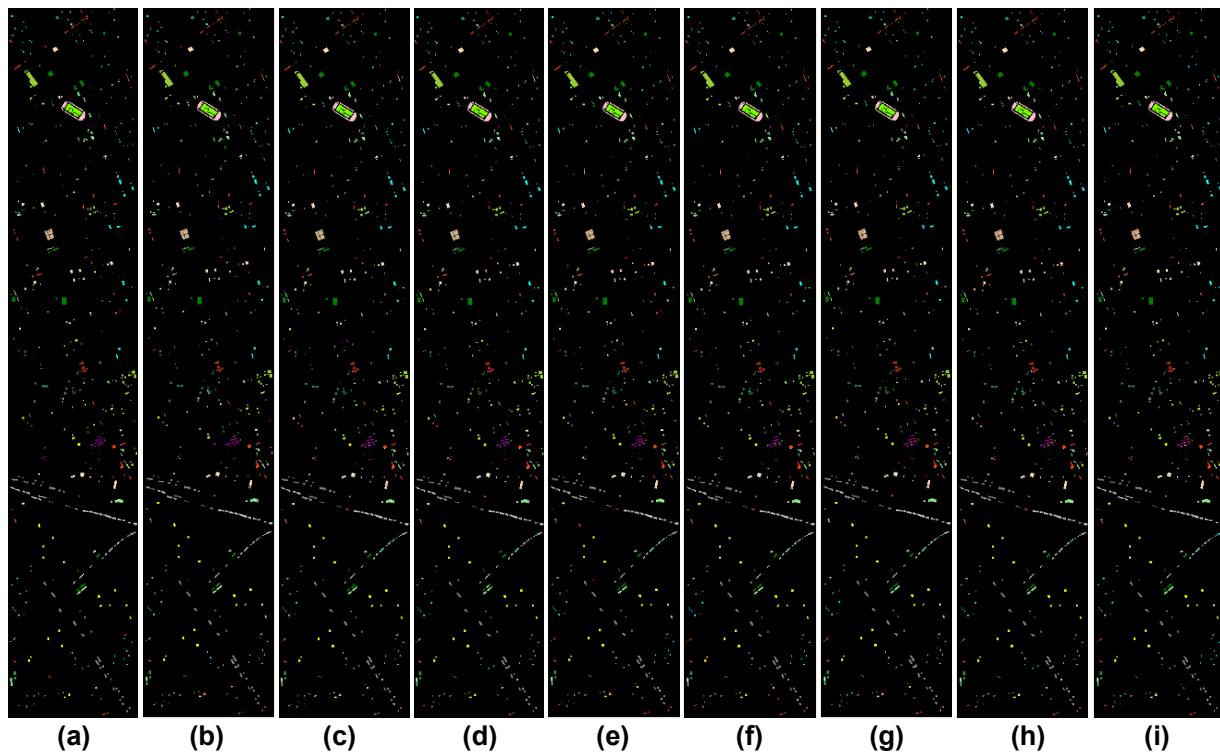


Fig. 14. (a) Ground Truth Image, Classification Maps of (b) SCMK (c) ULBP-SPG (d) R2MK (e) ASGSSK (f) ASMGSSK (g) SuperPCA (h) MsuperPCA (i) IAP for Houston 2013 dataset

spectral unmixing, dimensionality reduction, Band selection, Active learning, denoising, and anomaly detection. In this work, a unified post-processing framework for superpixels is presented. Also, a brief survey on various application areas of superpixels is provided. In spite of so many advantages, there exist several open challenges also in the implementation of superpixel segmentation. Determination of the number of superpixels and their position is a challenging task. There is also a need to develop dedicated superpixel evaluation metrics for hyperspectral images. Even though several superpixel segmentation algorithms exist in the literature, most of them are developed for RGB images. These algorithms are often modified before applying them to HSI. There is a need to develop dedicated superpixel segmentation algorithms specific to HSI. Also, deep learning techniques may be incorporated for generating superpixel segments in the future for HSI.

## REFERENCES

- [1] David R Thompson, Benjamin J Bornstein, Steve A Chien, Steven Schaffer, Daniel Tran, Brian D Bue, Rebecca Castaño, Damnhait F Gleeson, and Aaron Noell. Autonomous spectral discovery and mapping onboard the eo-1 spacecraft. *IEEE Transactions on Geoscience and Remote Sensing*, 51(6):3567–3579, 2012.
- [2] Woo-Yong Jang, Zahyun Ku, Augustine Urbas, John Derov, and Michael J Noyola. Plasmonic superpixel sensor for compressive spectral sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3471–3480, 2015.
- [3] Yao Qin, Lorenzo Bruzzone, and Biao Li. Learning discriminative embedding for hyperspectral image clustering based on set-to-set and sample-to-sample distances. *IEEE Transactions on Geoscience and Remote Sensing*, 58(1):473–485, 2019.
- [4] Li Ma, Melba M Crawford, and Jinwen Tian. Local manifold learning-based  $k$ -nearest-neighbor for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4099–4109, 2010.
- [5] Jon Atli Benediktsson, Philip H Swain, and Okan K Ersoy. Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data. *International Journal of Remote Sensing*, 14(15):2883–2903, 1993.
- [6] Pedram Ghamisi, Javier Plaza, Yushi Chen, Jun Li, and Antonio J Plaza. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geoscience and Remote Sensing Magazine*, 5(1):8–32, 2017.
- [7] Pedram Ghamisi, Emmanuel Maggiori, Shutao Li, Roberto Souza, Yuliya Tarabalka, Gabriele Moser, Andrea De Giorgi, Leyuan Fang, Yushi Chen, Mingmin 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 geoscience and remote sensing magazine*, 6(3):10–43, 2018.
- [8] Mathieu Fauvel, Yuliya Tarabalka, Jon Atli Benediktsson, Jocelyn Chanussot, and James C Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2012.
- [9] Le Sun, Chenyang Ma, Yunjie Chen, Hiuk Jae Shim, Zebin Wu, and Byeungwoo Jeon. Adjacent superpixel-based multiscale spatial-spectral kernel for hyperspectral classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(6):1905–1919, 2019.
- [10] Sen Jia, Bin Deng, Jiasong Zhu, Xiuping Jia, and Qingquan Li. Local binary pattern-based hyperspectral image classification with superpixel guidance. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):749–759, 2017.
- [11] Xiaofeng Ren and Jitendra Malik. Learning a classification model for segmentation. In *null*, page 10. IEEE, 2003.
- [12] David Stutz, Alexander Hermans, and Bastian Leibe. Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding*, 166:1–27, 2018.
- [13] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-

- of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.
- [14] Xinyu Wang, Yanfei Zhong, Liangpei Zhang, and Yanyan Xu. Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6287–6304, 2017.
- [15] Junjun Jiang, Jiayi Ma, Chen Chen, Zhongyuan Wang, Zhihua Cai, and Lizhe Wang. Superpca: A superpixelwise pca approach for unsupervised feature extraction of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8):4581–4593, 2018.
- [16] Chen Yang, Lorenzo Bruzzone, Haishi Zhao, Yulei Tan, and Renchu Guan. Superpixel-based unsupervised band selection for classification of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 56(12):7230–7245, 2018.
- [17] Zhaohui Xue, Shaoguang Zhou, and Pengfei Zhao. Active learning improved by neighborhoods and superpixels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 15(3):469–473, 2018.
- [18] Le Sun, Byeungwoo Jeon, Bushra Naz Soomro, Yuhui Zheng, Zebin Wu, and Liang Xiao. Fast superpixel based subspace low rank learning method for hyperspectral denoising. *IEEE Access*, 6:12031–12043, 2018.
- [19] Lang Ren, Liaoying Zhao, and Yulei Wang. A superpixel-based dual window rx for hyperspectral anomaly detection. *IEEE Geoscience and Remote Sensing Letters*, 2019.
- [20] Tanu Priya, Saurabh Prasad, and Hao Wu. Superpixels for spatially reinforced bayesian classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 12(5):1071–1075, 2015.
- [21] Yuxiang Zhang, Kang Liu, Yanni Dong, Ke Wu, and Xiangyun Hu. Semisupervised classification based on slic segmentation for hyperspectral image. *IEEE Geoscience and Remote Sensing Letters*, 2019.
- [22] Zhengqin Li and Jiansheng Chen. Superpixel segmentation using linear spectral clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1356–1363, 2015.
- [23] Jian Yao, Marko Boben, Sanja Fidler, and Raquel Urtasun. Real-time coarse-to-fine topologically preserving segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2947–2955, 2015.
- [24] Tianzhu Liu, Yanfeng Gu, Jocelyn Chanussot, and Mauro Dalla Mura. Multimorphological superpixel model for hyperspectral image classification. *IEEE Transactions on geoscience and remote sensing*, 55(12):6950–6963, 2017.
- [25] Gholamreza Akbarizadeh. A new statistical-based kurtosis wavelet energy feature for texture recognition of sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4358–4368, 2012.
- [26] Zeinab Tirandaz and Gholamreza Akbarizadeh. A two-phase algorithm based on kurtosis curvelet energy and unsupervised spectral regression for segmentation of sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(3):1244–1264, 2015.
- [27] Leyuan Fang, Nanjun He, Shutao Li, Pedram Ghamisi, and Jón Atli Benediktsson. Extinction profiles fusion for hyperspectral images classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(3):1803–1815, 2017.
- [28] Shanshan Li, Li Ni, Xiuping Jia, Lianru Gao, Bing Zhang, and Man Peng. Multi-scale superpixel spectral-spatial classification of hyperspectral images. *International journal of remote sensing*, 37(20):4905–4922, 2016.
- [29] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2):167–181, 2004.
- [30] Alastair P Moore, Simon JD Prince, Jonathan Warrell, Umar Mohammed, and Graham Jones. Superpixel lattices. In *2008 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008.
- [31] Olga Veksler, Yuri Boykov, and Paria Mehrani. Superpixels and supervoxels in an energy optimization framework. In *European conference on Computer vision*, pages 211–224. Springer, 2010.
- [32] Ming-Yu Liu, Oncel Tuzel, Srikumar Ramalingam, and Rama Chellappa. Entropy rate superpixel segmentation. In *CVPR 2011*, pages 2097–2104. IEEE, 2011.
- [33] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.
- [34] Andrea Vedaldi and Stefano Soatto. Quick shift and kernel methods for mode seeking. In *European conference on computer vision*, pages 705–718. Springer, 2008.
- [35] Luc Vincent and Pierre Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):583–598, 1991.
- [36] Alex Levinstein, Adrian Stere, Kiriakos N Kutulakos, David J Fleet, Sven J Dickinson, and Kaleem Siddiqi. Turbopixels: Fast superpixels using geometric flows. *IEEE transactions on pattern analysis and machine intelligence*, 31(12):2290–2297, 2009.
- [37] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels. Technical report, 2010.
- [38] Xin Huang and Liangpei Zhang. An svm ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE transactions on geoscience and remote sensing*, 51(1):257–272, 2012.
- [39] Sen Jia, Kuilin Wu, Jiasong Zhu, and Xiuping Jia. Spectral-spatial gabor surface feature fusion approach for hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):1142–1154, 2018.
- [40] He Yang, Qian Du, and Ben Ma. Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 7(4):875–879, 2010.
- [41] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2):415–425, 2002.
- [42] Ting Lu, Shutao Li, Leyuan Fang, Xiuping Jia, and Jón Atli Benediktsson. From subpixel to superpixel: A novel fusion framework for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8):4398–4411, 2017.
- [43] Shutao Li, Ting Lu, Leyuan Fang, Xiuping Jia, and Jon Atli Benediktsson. Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7416–7430, 2016.
- [44] Ting Lu, Shutao Li, Leyuan Fang, Lorenzo Bruzzone, and Jón Atli Benediktsson. Set-to-set distance-based spectral-spatial classification of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7122–7134, 2016.
- [45] Xianghai Cao, Hongxia Lu, Meiru Ren, and Licheng Jiao. Non-overlapping classification of hyperspectral imagery with superpixel segmentation. *Applied Soft Computing*, 83:105630, 2019.
- [46] Leyuan Fang, Shutao Li, Wuhui Duan, Jinchang Ren, and Jón Atli Benediktsson. Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels. *IEEE transactions on geoscience and remote sensing*, 53(12):6663–6674, 2015.
- [47] Philip Sellars, Angelica Aviles-Rivero, and Carola-Bibiane Schönlieb. Superpixel contracted graph-based learning for hyperspectral image classification. *arXiv preprint arXiv:1903.06548*, 2019.
- [48] Tugcan Dundar and Taner Ince. Sparse representation-based hyperspectral image classification using multiscale superpixels and guided filter. *IEEE Geoscience and Remote Sensing Letters*, 16(2):246–250, 2018.
- [49] Jiayi Li, Hongyan Zhang, and Liangpei Zhang. Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(10):5338–5351, 2015.
- [50] Yong Ma, Chang Li, Hao Li, Xiaoguang Mei, and Jiayi Ma. Hyperspectral image classification with discriminative kernel collaborative representation and tikhonov regularization. *IEEE Geoscience and Remote Sensing Letters*, 15(4):587–591, 2018.
- [51] Qian Liu, Zebin Wu, Le Sun, Yang Xu, Lu Du, and Zhihui Wei. Kernel low-rank representation based on local similarity for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(6):1920–1932, 2019.
- [52] Fan Fan, Yong Ma, Chang Li, Xiaoguang Mei, Jun Huang, and Jiayi Ma. Hyperspectral image denoising with superpixel segmentation and low-rank representation. *Information Sciences*, 397:48–68, 2017.
- [53] Masoumeh Rahmani and Gholamreza Akbarizadeh. Unsupervised feature learning based on sparse coding and spectral clustering for segmentation of synthetic aperture radar images. *IET Computer Vision*, 9(5):629–638, 2015.
- [54] Chih-Jen Lin. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10):2756–2779, 2007.
- [55] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 7(6):2094–2107, 2014.

- [56] Yushi Chen, Xing Zhao, and Xiuping Jia. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381–2392, 2015.
- [57] Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251, 2016.
- [58] Lichao Mou, Pedram Ghamisi, and Xiao Xiang Zhu. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3639–3655, 2017.
- [59] Lin Zhu, Yushi Chen, Pedram Ghamisi, and Jón Atli Benediktsson. Generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9):5046–5063, 2018.
- [60] Xiaorui Ma, Hongyu Wang, and Jie Geng. Spectral-spatial classification of hyperspectral image based on deep auto-encoder. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9):4073–4085, 2016.
- [61] Jun Yue, Wenzhi Zhao, Shanjun Mao, and Hui Liu. Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*, 6(6):468–477, 2015.
- [62] Wei Li, Guodong Wu, Fan Zhang, and Qian Du. Hyperspectral image classification using deep pixel-pair features. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):844–853, 2016.
- [63] Yazhou Liu, Guo Cao, Quansen Sun, and Mel Siegel. Hyperspectral classification via deep networks and superpixel segmentation. *International Journal of Remote Sensing*, 36(13):3459–3482, 2015.
- [64] Li Liu, Yuebin Wang, Junhuan Peng, Liqiang Zhang, Bing Zhang, and Yun Cao. Latent relationship guided stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5):3711–3725, 2020.
- [65] Cheng Shi and Chi-Man Pun. Multiscale superpixel-based hyperspectral image classification using recurrent neural networks with stacked autoencoders. *IEEE Transactions on Multimedia*, 2019.
- [66] Yuebin Wang, Jie Mei, Liqiang Zhang, Bing Zhang, Anjian Li, Yibo Zheng, and Panpan Zhu. Self-supervised low-rank representation (sslr) for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10):5658–5672, 2018.
- [67] Jie Feng, Liguang Liu, Xianghai Cao, Licheng Jiao, Tao Sun, and Xiangrong Zhang. Marginal stacked autoencoder with adaptively-spatial regularization for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(9):3297–3311, 2018.
- [68] Sen Jia, Bin Deng, Jiasong Zhu, Xiuping Jia, and Qingquan Li. Superpixel-based multitask learning framework for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5):2575–2588, 2017.
- [69] Jianjun Liu, Zebin Wu, Zhiyong Xiao, and Jinlong Yang. Region-based relaxed multiple kernel collaborative representation for hyperspectral image classification. *IEEE Access*, 5:20921–20933, 2017.
- [70] Haoyang Yu, Lianru Gao, Wenzhi Liao, Bing Zhang, Aleksandra Pižurica, and Wilfried Philips. Multiscale superpixel-level subspace-based support vector machines for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 14(11):2142–2146, 2017.
- [71] Binge Cui, Xiaoyun Xie, Xiudan Ma, Guangbo Ren, and Yi Ma. Superpixel-based extended random walker for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(6):3233–3243, 2018.
- [72] Jie Mei, Yuebin Wang, Liqiang Zhang, Bing Zhang, Suhong Liu, Panpan Zhu, and Yingchao Ren. Psal: Pixel-level and superpixel-level aware subspace learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):4278–4293, 2019.
- [73] Puhong Duan, Xudong Kang, Shutao Li, Pedram Ghamisi, and Jón Atli Benediktsson. Fusion of multiple edge-preserving operations for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):10336–10349, 2019.
- [74] Zhikun Chen, Junjun Jiang, Chong Zhou, Shaoyuan Fu, and Zhihua Cai. Superbf: Superpixel-based bilateral filtering algorithm and its application in feature extraction of hyperspectral images. *IEEE Access*, 7:147796–147807, 2019.
- [75] Sen Jia, Zhijie Lin, Bin Deng, Jiasong Zhu, and Qingquan Li. Cascade superpixel regularized gabor feature fusion for hyperspectral image classification. *IEEE transactions on neural networks and learning systems*, 2019.
- [76] Sen Jia, Xianglong Deng, Jiasong Zhu, Meng Xu, Jun Zhou, and Xiuping Jia. Collaborative representation-based multiscale superpixel fusion for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10):7770–7784, 2019.
- [77] Xiaowei Ji, Ying Cui, Heng Wang, Long Teng, Lingxiu Wang, and Liguang Wang. Semisupervised hyperspectral image classification using spatial-spectral information and landscape features. *IEEE Access*, 7:146675–146692, 2019.
- [78] Danfeng Hong, Xin Wu, Pedram Ghamisi, Jocelyn Chanussot, Naoto Yokoya, and Xiao Xiang Zhu. Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification. *arXiv preprint arXiv:1912.08847*, 2019.
- [79] Wei Huang, Yao Huang, Hua Wang, Yan Liu, and Hiuk Jae Shim. Local binary patterns and superpixel-based multiple kernels for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:4550–4563, 2020.
- [80] Qingming Leng, Haiou Yang, Junjun Jiang, and Qi Tian. Adaptive multiscale segmentations for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2020.
- [81] Leyuan Fang, Shutao Li, Xudong Kang, and Jón Atli Benediktsson. Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model. *IEEE Transactions on Geoscience and Remote Sensing*, 53(8):4186–4201, 2015.
- [82] Shuzhen Zhang, Shutao Li, Wei Fu, and Leiyan Fang. Multiscale superpixel-based sparse representation for hyperspectral image classification. *Remote Sensing*, 9(2):139, 2017.
- [83] Bing Tu, Chengle Zhou, Xiaolong Liao, Guoyun Zhang, and Yishu Peng. Spectral-spatial hyperspectral classification via structural-kernel collaborative representation. *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [84] Jinhuan Xu, James E Fowler, and Liang Xiao. Hypergraph-regularized low-rank subspace clustering using superpixels for unsupervised spatial-spectral hyperspectral classification. *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [85] Ge Li, Lingling Li, Hao Zhu, Xu Liu, and Licheng Jiao. Adaptive multiscale deep fusion residual network for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):8506–8521, 2019.
- [86] Min Wang, Bowen Zhang, Xi Pan, and Shuyuan Yang. Group low-rank nonnegative matrix factorization with semantic regularizer for hyperspectral unmixing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4):1022–1029, 2018.
- [87] Ricardo Augusto Borsoi, Tales Imbiriba, José Carlos Moreira Bermudez, and Cédric Richard. A fast multiscale spatial regularization for sparse hyperspectral unmixing. *IEEE Geoscience and Remote Sensing Letters*, 16(4):598–602, 2018.
- [88] Longshan Yang, Junhuan Peng, Huiwei Su, Linlin Xu, Yuebin Wang, and Bowen Yu. Combined nonlocal spatial information and spatial group sparsity in nmf for hyperspectral unmixing. *IEEE Geoscience and Remote Sensing Letters*, 2019.
- [89] Ricardo Augusto Borsoi, Tales Imbiriba, and José Carlos Moreira Bermudez. A data dependent multiscale model for hyperspectral unmixing with spectral variability. *arXiv preprint arXiv:1808.01047*, 2018.
- [90] Hao Li, Ruiyi Feng, Lizhe Wang, Yanfei Zhong, and Liangpei Zhang. Superpixel-based reweighted low-rank and total variation sparse unmixing for hyperspectral remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2020.
- [91] Souvick Mukherjee, Minshan Cui, and Saurabh Prasad. Spatially constrained semisupervised local angular discriminant analysis for hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4):1203–1212, 2017.
- [92] Renlong Hang and Qingshan Liu. Dimensionality reduction of hyperspectral image using spatial regularized local graph discriminant embedding. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(9):3262–3271, 2018.
- [93] Behnam Asghari Beirami and Mehdi Mokhtarzade. Band grouping superpca for feature extraction and extended morphological profile production from hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [94] Yulei Tan, Laijun Lu, Lorenzo Bruzzone, Renchu Guan, Zhiyong Chang, and Chen Yang. Hyperspectral band selection for lithologic discrimination and geological mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020.
- [95] Jieliang Guo, Xiong Zhou, Jun Li, Antonio Plaza, and Saurabh Prasad. Superpixel-based active learning and online feature importance learning

for hyperspectral image analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(1):347–359, 2016.

[96] Xiong Zhou and Saurabh Prasad. Active and semisupervised learning with morphological component analysis for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 14(8):1348–1352, 2017.

[97] Chenyong Liu, Jun Li, and Lin He. Superpixel-based semisupervised active learning for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(1):357–370, 2018.

[98] Chengyong Zheng, Ningning Wang, and Jing Cui. Hyperspectral image classification with small training sample size using superpixel-guided training sample enlargement. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10):7307–7316, 2019.

[99] Junjun Jiang, Jiayi Ma, Zheng Wang, Chen Chen, and Xianming Liu. Hyperspectral image classification in the presence of noisy labels. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):851–865, 2018.

[100] Bing Tu, Chengle Zhou, Danbing He, Siyuan Huang, and Antonio Plaza. Hyperspectral classification with noisy label detection via superpixel-to-pixel weighting distance. *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[101] Zhihong Huang and Shutao Li. From difference to similarity: A manifold ranking-based hyperspectral anomaly detection framework. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10):8118–8130, 2019.

[102] Zhihong Huang, Xudong Kang, Shutao Li, and Qiaobo Hao. Game theory-based hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[103] Yanan Gao, Tongkai Cheng, and Bin Wang. Nonlinear anomaly detection based on spectral-spatial composite kernel for hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 2020.

[104] Zhihong Huang, Leyuan Fang, and Shutao Li. Subpixel-pixel-superpixel guided fusion for hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[105] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012.

[106] Melba M Crawford, Devis Tuia, and Hsiuhan Lexie Yang. Active learning: Any value for classification of remotely sensed data? *Proceedings of the IEEE*, 101(3):593–608, 2013.

[107] Devis Tuia, Michele Volpi, Loris Copa, Mikhail Kanevski, and Jordi Muñoz-Mari. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):606–617, 2011.

[108] Qiangqiang Yuan, Liangpei Zhang, and Huanfeng Shen. Hyperspectral image denoising employing a spectral-spatial adaptive total variation model. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10):3660–3677, 2012.

[109] Tao Lin and Salah Bourennane. Survey of hyperspectral image denoising methods based on tensor decompositions. *EURASIP journal on Advances in Signal Processing*, 2013(1):186, 2013.

[110] Tao Lin and Salah Bourennane. Hyperspectral image processing by jointly filtering wavelet component tensor. *IEEE Transactions on Geoscience and Remote Sensing*, 51(6):3529–3541, 2012.

[111] Behnood Rasti, Johannes R Sveinsson, and Magnus Orn Ulfarsson. Wavelet-based sparse reduced-rank regression for hyperspectral image restoration. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10):6688–6698, 2014.

[112] Forough Sharifzadeh, Gholamreza Akbarizadeh, and Yousef Seifi Kaviani. Ship classification in sar images using a new hybrid cnn-mlp classifier. *Journal of the Indian Society of Remote Sensing*, 47(4):551–562, 2019.

[113] Bing Tu, Xianchang Yang, Nanying Li, Xianfeng Ou, and Wei He. Hyperspectral image classification via superpixel correlation coefficient representation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(11):4113–4127, 2018.

[114] Tianming Zhan, Zhenyu Lu, Minghua Wan, and Guowei Yang. Multi-scale superpixel kernel-based low-rank representation for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 2019.

[115] David R Martin, Charless C Fowlkes, and Jitendra Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE transactions on pattern analysis and machine intelligence*, 26(5):530–549, 2004.

[116] Alexander Schick, Mika Fischer, and Rainer Stiefelwagen. Measuring and evaluating the compactness of superpixels. In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, pages 930–934. IEEE, 2012.



Deep Learning, and Remote sensing.

**Subhashree Subudhi** received the B.Tech degree in Electronics and Communication Engineering from Biju Patnaik University of Technology, Bhubaneswar, India in 2012 and M.Tech degree in VLSI and Embedded System Design with Siksha 'O' Anusandhan University, Bhubaneswar, India in 2015. She is currently pursuing the PhD degree in electronics and communication engineering with the International Institute of Information Technology, Bhubaneswar, India. Her research interests include Hyperspectral Image Processing, Machine Learning,



Image Processing, Band

**Ram Narayan Patro** received the B.Tech degree in Electronics and Communication Engineering from Biju Patnaik University of Technology, Bhubaneswar, India in 2009 and M.Tech degree in VLSI and Embedded System Design with Jawaharlal Nehru Technological University, Hyderabad, India in 2015. He is currently pursuing the PhD degree in electronics and communication engineering with the International Institute of Information Technology, Bhubaneswar, India. His research interests includes Embedded System, Signal processing, Hyperspectral Selection, Machine Learning and Remote sensing.



associate member of IEI (India).

**Pradyut Kumar Biswal** (M'11; SM'20) obtained his M.Tech and Ph.D. from Visvesvaraya National Institute of Technology, Nagpur and Indian Institute of Technology, Kharagpur in 2002 and 2011 respectively. Currently, he is positioned as an Assistant Professor in International Institute of Information Technology, Bhubaneswar, India. His research interest includes VLSI Architecture design for signal and image processing algorithms, Biomedical signal and image processing, and Hyperspectral image processing. He is a Senior Member of IEEE, USA and



**Fabio Dell'Acqua** (M'01; SM '07) is Full Professor of Remote Sensing at the Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy. He teaches courses in remote sensing at the University of Pavia, and occasionally at other institutions. His research interests focus on radar data processing and radar/optical data fusion for risk-related applications. In this area he participated in several research projects, both at national and international level. From 2011 to 2015, he organized yearly editions of an International Summer School

on Data Fusion in Aerospace Applications, which attracted up to 40 students from around the world. In 2014, he co-founded a university spin-off company, named Ticinum Aerospace, to exploit commercially his research results in the use of EO data for risk management. He is a Life Member of the Technical and Scientific Board of the Lombardy Aerospace Industry Cluster.

APPENDIX  
GENERAL NOTATIONS

The general notations used in the paper are provided in Table. V.

TABLE V  
GENERAL NOTATIONS

Symbol	Meaning
$K$	number of superpixels
$\mathcal{G}$	Graph
$V$	Vertex
$E$	Edge
$w$	weight
$\mathcal{H}$	entropy rate function
$E'$	selected edge set
$P$	probability
$\mathfrak{B}$	balancing function
$Z$	cluster membership distribution
$\mathcal{N}$	number of connected components
$S$	graph partitioning for edge set / Superpixel map
$\lambda$	weight of balancing term
$H$	input HSI
$n$	pixels
$B$	spectral bands
$Q$	window size
$\mathfrak{D}$	distance
$(r, u)$	location of pixel $i$
$C$	classes
$\mathfrak{J}$	total training samples for all classes
$T$	number of features
$\mathcal{P}$	predicted label of test sample
$\delta$	indicator function
$h$	predefined scalar
$D$	structured dictionary
$N$	number of samples in $D_K$
$X$	each pixel of HSI, $X \in \mathbb{R}^{B \times 1}$
$L$	sparsity level
$A$	sparse coefficient matrix
Err	Error
$\mathfrak{A}$	pixel weight set
$\mathfrak{B}$	collaborative coefficient set
$H'$	clean HSI
$E_{sparse}$	sparse error term
$N_{gaussian}$	Gaussian noise
$M$	endmembers
$A$	abundances
$\vartheta_p$	spatial superpixel group
$c$	superpixel confidence index
$W$	superpixel-wise weight matrix
$G$	Ground truth
$FN$	False positive
$TN$	True negative
$UE$	Under-segmentation error
$F$	image size in pixels
$EV$	Explained Variance
$\mu$	mean
$I$	image
$CO$	compactness
$Ar$	Area
$Peri$	Perimeter